



Online handwriting recognition systems for Indic and non-Indic scripts: a review

Harjeet Singh¹ · R. K. Sharma² · V. P. Singh²

Published online: 8 August 2020
© Springer Nature B.V. 2020

Abstract

Handwriting recognition is one of the challenging tasks in the area of pattern recognition and machine learning. Handwriting recognition has two flavors, namely, Offline Handwriting Recognition and Online Handwriting Recognition. Though, saturation level has been achieved in machine printed (Offline) character recognition. Presently, due to dramatical development in IT sector, touch-based devices are available in the market with efficient processing capabilities. With this revolution, research in the area of handwriting recognition has become more popular in real-time (Online) mode. In this paper, a comprehensive review has been reported for online handwriting recognition of non-Indic and Indic scripts. The six non-Indic-scripts and eight Indic script namely, Arabic, Chinese, Japanese, Persian, Roman, Thai, and, Assamese, Bangla, Devanagari, Gurmukhi, Kannada, Malayalam, Tamil, Telugu, respectively have been considered in this article. This study comprises introduction of online handwriting recognition process, various challenges, motivations, feature extraction, and classification methodologies, used for recognizing the various scripting languages. Moreover, an effort has been made to provide the list of publicly available online handwritten dataset for various scripting languages. This study also provides the recognition and beneficial assistance to the novice researchers in field of handwriting recognition by providing a nut shell studies of various feature extraction strategies and classification techniques, used for the recognition of both Indic and non-Indic scripts.

Keywords Online handwriting recognition · Indic scripts · non-Indic scripts · Pre-processing · Feature extraction · Classification · Datasets · Post-processing

✉ Harjeet Singh
harjeet.research@gmail.com; harjeet.singh@chitkara.edu.in

R. K. Sharma
rksharma@thapar.edu

V. P. Singh
vpsingh@thapar.edu

¹ Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India

² Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab, India

1 Introduction

Handwriting is one of the classical modes of communication between human beings. Prior to handwriting, verbal communication and sign-language were two basic ways of communication. Handwriting helps in recording the day-to-day activities, culture, mathematics, literature and much more. Handwriting can be referred as a system of standard symbols, where combinations of these symbols convey some meaningful communication. Growth in culture and civilization gave birth to languages, scripts, and symbols to fulfill the need for better communication. Handwriting is a unique property of every individual and writing style varies from person to person. The common causes of variation include, state of mind, mood of the person, writing environment, writing medium etc.

Due to rapidly growth in Information Technology (IT) sector, the human-computer-interfacing devices (i.e., touch-based phones, Tablet-PCs) based online handwriting recognition has become more popular. In online handwriting recognition, the trajectories of stylus/digital pen movements are recorded and analyze to identify the handwritten symbol, while in offline handwriting recognition the handwritten images of characters/words are processed and classified into different classes. In both the cases (offline and online), handwritten data is processed through several phases (i.e., size-normalization, interpolation, smoothing, re-sampling, feature extraction, recognition etc.) in order to get its digital form. Online handwriting deals with spatio-temporal features of the input, whereas the offline handwriting deals with spatio-luminance analysis of the input images (Parui et al. 2008).

The primary objective of this paper is to present a comprehensive review of the state of the art in online handwriting recognition systems for Indic and non-Indic scripts. In this connection, eight Indic-scripts, namely, Assamese, Bangla, Devanagari, Gurmukhi, Kannada, Malayalam, Tamil, and Telugu; and six non-Indic-scripts, namely, Arabic, Chinese, Japanese, Persian, Roman, and Thai have been reviewed. In this review, a number of aspects i.e., writing structure of script, pre-processing, feature extraction, dataset, post-processing, and experimental results have been reported. of online handwriting recognition process including the complexity levels have been discussed.

This paper has been organized in seven sections. In Sect. 1, we have discussed the online handwriting recognition process, various challenges, motivation and methodologies used for the recognition of scripting languages. Section 2 presents the research work done for the recognition of non-Indic scripts. In Sect. 3, the research work related to the Indian languages have been discussed. In the end of this section, script-wise summarization of research work for non-Indic and Indic script have been discussed by considering some important parameters such as preprocessing, features used, methodology/tool used and accuracy achieved. Comparative study of different classification techniques have been discussed in Sect. 4. Moreover, in Sect. 5, a list of publicly available online handwritten dataset of various Indic and non-Indic script have been reported. Future scope of the present study has been explained in Sect. 6. Finally, Sect. 7 summarizes the work done in the present study.

1.1 Online handwriting recognition process

In general, online handwriting recognition process is the composition of five phases, namely, data collection phase, data pre-processing phase, feature extraction phase, recognition phase, and post-processing phase. It is worth mentioning here that, among these

phases, Data collection is the primary and common step in all the online handwriting recognition systems, developed for Indic and non-Indic scripts. In this phase, handwritten samples are collected with the help of touch based digital devices include, Tablet-PCs, smart-phones, digitizers, etc. The development and implementation of the remaining phases is vary among different scripts. Figure 1 illustrate the common process of a online handwriting recognition system, followed by every scripting language.

1.2 Challenges and motivations

Online handwriting recognition systems have their own challenges. These challenges are: (1) writers specific, (2) machine specific, and (3) scripting language specific. Following are some common challenges, arise in almost all the scripting languages.

- (a) *Variation in handwriting* Handwriting is a free-form activity, and there are many ways to write even the simplest character. Different writers write the same character with different combination of strokes and with different size of strokes. These variations are observed geometrically. Some of the common geometrical properties are position, size, and aspect ratio of a stroke.
- (b) *Constrained and unconstrained handwriting* Handwriting can be classified into two categories: constrained and unconstrained handwriting. In constrained handwriting, a writer writes the characters/aksharas with some given restriction i.e. written character within the defined box (isolated), space-discrete handwriting, where each character is written separately and herein adjacent character does not touch each other, run-on discrete handwriting, where separately written characters may touch each other etc. On the other side, in unconstrained handwriting, writers have no such restriction to write characters/word. Mixture of space-discrete and cursive handwriting styles comes in this category. All these categories are illustrated using the online handwritten Gurmukhi word "Patiala", (name of a district in Punjab State, India) in Fig. 2.
- (c) *Behavior, personal and hardware factor* Human beings are emotional by nature. The handwriting style is influenced by the emotions of a writer. Behavior describes the

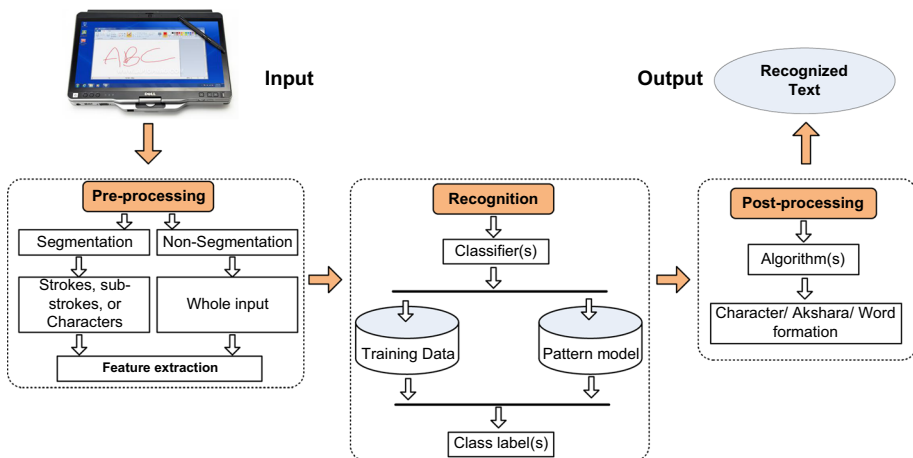
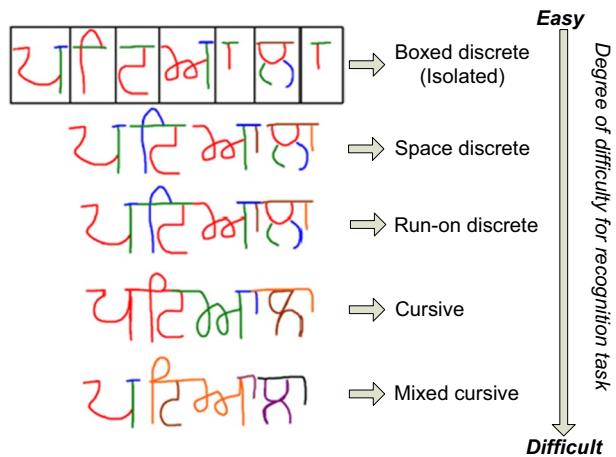


Fig. 1 Flow chart of the online handwriting recognition system

Fig. 2 Types of different handwriting styles (different colors indicate different stroke within a word sample)



way of presenting the handwriting, that could be stressful, flurry, excitement, sad, or distraction (Wing 1979). The personal factors include the writer's handedness, either left-handed or right-handed. It is observed that in both the types, writers use different positions and directions while writing. In online handwriting recognition, hardware is a major factor, which affects the performance of handwriting recognition. The screen size and resolutions of a hardware device may also affect the handwriting style.

- (d) *Writer dependent and writer-independent recognition systems* The writer-dependent recognition system is limited to recognize the handwriting styles of a specific writer. Generally, the writer-dependent recognition system is trained with the handwriting patterns of a known writer whose handwriting will be recognized in future. On the other hand, the writer-independent recognition system is trained with the handwriting patterns of unknown writers. Here, all the possible and commonly used style variations of different writers are considered for training the system. Therefore, it is a challenging task to train a recognition system with large number of different writers' handwriting samples.
- (e) *Complex structure of a scripting language* Every scripting language has their own properties, characteristics, and writing structure. Structure includes, composition of a script using *Consonants*, *Vowel*, *Conjuncts*, cursive writing style etc. Apart from these, some scripts are written in three horizontal zones (Belhe et al. 2012; Samanta et al. 2014; Singh et al. 2018a). Therefore, in online handwriting recognition systems, structure of a particular script also a major challenge.

Due to the rapidly growth in Information Technology (IT) sector, the demand of smart phones-, Tablet-PCs-, digitizers-based application has been increased. The motivation here is to utilize these devices by developing the online handwriting recognition based applications and overcome the aforementioned challenges.

1.3 Methodologies

In order to recognize the online handwritten text, a classifier is trained by using the extracted features. After training the classifier, this trained model is used further to recognize the handwritten text. The process of classifying a class for unknown observation

from the trained model is a indicative problem of machine learning. The most common statistical, structural, syntactical, and neural network based classification methods that are used for handwriting recognition are: Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Hidden Markov Models (HMMs), Convolution Neural Networks (CNNs), and Elastic Matching. These methodologies are described as below and summarized in Table 1.

1.3.1 Statistical methods

The application of statistical methods extract information from data and provide different ways to access the robustness of research outputs. Statistical methods can be contrasted with deterministic methods, which are appropriate where observations are exactly reproduceable or are assumed to be so. Statistical methods are probabilities based methods and categorized into two types, namely, parametric and non-parametric methods. In the parametric methods, the samples of handwriting are considered as statistical variables from the distribution that are characterized by a set of parameters and each class is characterized by its own set of parameters. On the basis of training data these parameters are selected. Hidden Markov Model (HMM) classifier is an example of parametric statistical methods. In contrast, the non-parametric methods make use of the training data to estimate the value(s) of potentially unknown parameters. The complexity of non-parametric methods increases with the increase of training dataset. One of the common non-parameteric method based classifier is k -Nearest Neighbor (k -NN). The parametric methods are widely preferred as compared to non-parametric methods in term of easy computation.

1.3.2 Structural methods

In online handwriting recognition, the structural methods deal with the recognition of handwriting patterns via elastic matching of strings, graphs, or other structural description. The topological shape of the character pattern or strokes sequence and the x -, y -coordinates in the strokes are recorded as structural representation and resemble well to such a mechanism like human perception. During classification, the structure of online handwritten captured pattern is matched with the structural templates of all the defined classes and is classified as a class template, which has minimum distance and maximum similarity.

1.3.3 Syntactical methods

Syntactical methods deal with rules and grammar. Syntactical methods provide a description of construction of a pattern from the primitives. In syntactic pattern recognition, a formal analogy is drawn between the structure of patterns and the syntax of a language. The sentences are displayed as sentences belonging to a language, primitives are displayed as the alphabets of the language, and the sentences are generated according to a grammar. Hence, a large collection of the complex patterns can be described by a small number of primitives and grammatical rules. The grammar for each pattern class in this type of methods must be inferred from the available training samples (Jain et al. 2000).

Table 1 An illustration of Pattern Recognition (PR) methodologies

Methodology	Representation	Recognition function	Tool(s)/technique(s)	Evaluation criteria
Statistical	Features	Discriminant function	HMM, K-NN, SVM, and Language mode((s)	Classification error
Structural	Pixels, curves, sample shapes	Correlation, distance measure	Elastic matching	Classification error
Syntactical	Pattern/subpatterns (primitives)	Rules, grammar	Finite state automata (FSA)	Acceptance error
Neural network	Samples, pixels, features	Network function	ANN, MLP, RNN, DNN	Mean square error

1.3.4 Neural network based methods

Neural network methods are used for learning, generalization, adaptivity, fault tolerance with distributed representation, and computation. A neural network is a network of weighted graphs, where the nodes are artificial neurons and directed edges are connections between neuron input/output. The key feature of a neural network method is the ability to learn complex non-linear input/output relationships by using sequential training procedures. Feed forward neural network methods are the most commonly used for pattern classification tasks, which are composed of multi-layered perceptron and Radial-Basis Function (RBF) networks (Jain et al. 1996).

2 Work done for the recognition of non-Indic scripts

However, there are many non-Indic scripting languages, used for communication across the world. But in this article, the most common non-scripting languages, namely, Roman, Chinese, Japanese, Arabic, and Thai have been studied for the recognition of online handwriting recognition.

2.1 Arabic script

Arabic is a native language of Arab league nations. It has 313 million of native speakers across the world. Unlike Latin characters, the characters of the Arabic language are always written cursively from right to left direction. An Arabic word is written with one or more connected portions, and every portion has one or more characters that are not connected from the left side with succeeding character. Depending on the position within a connected portion of the word, every character has more than one shape. Hence, the recognition of Arabic language is quite complicated. Almuallim and Yamaguchi (1987) proposed a method to recognize the Arabic cursive handwriting. The method first pre-processes the handwritten word and further, the pre-processed word is segmented into strokes and afterwards, classification is employed at stroke-level. They have used geometrical and topological features for stroke classification. Eventually, after combining the recognized strokes, the final word is produced. They achieved a good recognition accuracy on a dataset of 400 words written by two persons. El-Wakil and Shoukry (1989) proposed an online handwritten isolated character recognition system for the Arabic language. They have used template matching with tree structure and K -NN classifier for the recognition. They tested the performance of the recognition systems with a dataset of 60 isolated Arabic characters. Their recognition system achieved an accuracy of 84.0% for characters and of 93.0% when they considered the features with manually assigned weights. Beigi et al. (1994) presented the challenges of handwriting recognition for Farsi, Arabic, and other languages with similar writing styles. In their work, they segmented the handwritten word information into strokes. The features they have employed include, stroke based extreme velocities and geometric features. They have performed the classification using HMM classifier. For building the training model, they have considered 600 samples each of 10 Arabic digits, written by 20 writers. In the testing phase, they have collected 5 samples of each digit from 14 new writers. Their writer independent recognizer produced 93.1% accuracy. Bouslama and Amin (1998) have presented a hybrid approach for the automatic recognition of

handwritten Arabic characters. The algorithm was based on features extracted by structural techniques and modeled by fuzzy sets. The features they have used in their work are: lines, curves, and diacritic points. They have used simple fuzzy if-then rules to classify Arabic characters. Alsallakh and Safadi (2006) presented an AraPen, a trainable Arabic handwriting recognition system with a high recognition rate for non-cursive characters. Their recognition process was based on mathematical matching techniques (based on the similarity between two strokes) and DTW distance. The pre-processing steps include smoothing, size normalization, and remsapling. The features they considered are: x - and y -coordinates points, tangent angles series, winding value, and aspect ratio. Their experiments achieved 91.0% accuracy with default pattern set and 98.0% accuracy after training the system. Biadsy et al. (2006) proposed a Hidden Markov Model (HMM) based system to recognize the Arabic script. For Arabic-word recognition, they have used word-part network, $WPN^*_{k,i}$ for $1 \leq i \leq k$, and the efficient Viterbi algorithm. El-Abed et al. (2009) presented a competition on Online Arabic handwriting recognition, wherein ADAB-database, consisting of 23,252 words was used, which were written by 132 writers. Later on, Kherallah et al. (2011) discussed that the online Arabic handwriting recognition system made a remarkable progress. In this competition, the most of the participants showed a very high accuracy and also a fast recognition speed. Moreover, it was demonstrated that HMM is also a powerful tool for handwriting recognition, and all the recognizers in this competition used HMM approach for classification. Mahmoud et al. (2018) presented the online handwritten Arabic text database, named as online-KHATT. This database is collected by 623 native writers. Some samples of the Arabic text lines are shown in Fig. 3.

2.2 Chinese script

Chinese language is the most popular language of Asia region. Nearly 1.2 billion people (around 16% of the world's population) speak Chinese. Chinese script's characters are classified into three categories, namely, Chinese characters, simplified Chinese characters, and Japanese Kanji. Liu et al. (2004) presented the advances in online Chinese character recognition with emphasis on the research work from the 1990s. The pre-processing used in their work include noise elimination, data reduction, and shape normalization. For pattern representation of input patterns and database modeling, they grouped the schemes into three categories such as statistical, structural, and hybrid statistical-structural. The character classification methodologies they have employed in their work

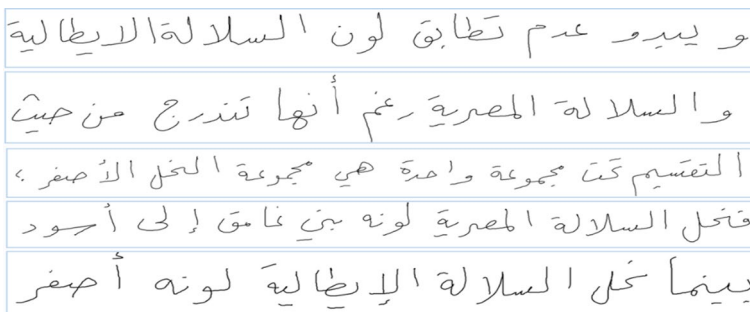
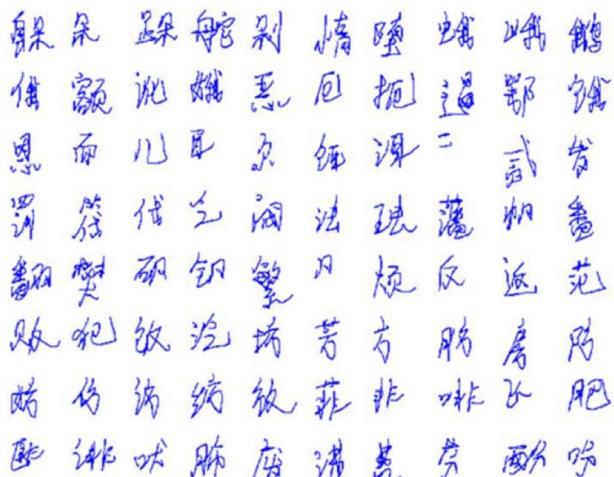


Fig. 3 Examples of online handwritten Arabic text lines

included structural matching, probabilistic matching (HMM), and statistical classification. The integration of linguistic information largely reduced the error rate in their work. The recognition performance for Chinese characters reported by them is 98.0% on regular scripts and 90.0% on fluent-regular scripts. Thereafter, Bai and Huo (2005) presented the online handwritten Chinese character recognition using 8-directional features. In this work, the pre-processing steps include linear size normalization, adding imaginary strokes, nonlinear shape normalization, equidistance resampling, and smoothing. Their pre-processing process produces a 64×64 normalized character sample. Thereafter, 8-directional features are extracted from each online trajectory point, and 8 directional pattern images are generated. Finally, a set of 512-dimensional feature vector is formed corresponding to each character class. The highest character recognition accuracy they have achieved in their experiments is 99.8%. Liu et al. (2013) have presented online and offline handwritten Chinese character recognition. They have worked on benchmarking new databases. The Institute of Automation of Chinese Academy of Sciences (CASIA) and National Laboratory of Pattern Recognition (NLPR) released the unconstrained online and offline Chinese handwriting databases OLHWDB and HWDB, which contain isolated character samples and handwritten texts produced by 1020 writers. They have presented their benchmarking results using state-of-the-art methods on the isolated character datasets OLHWDB1.0 and HWDB1.0 (referred as DB1.0), OLHWDB1.1 and HWDB1.1 (referred as DB1.1). A few samples of this database are shown in Fig. 4.

The DB1.1 covers 3755 Chinese character classes. In the preprocessing phase, they have used 1-D and pseudo 2-D normalization methods for evaluation. In the features extraction phase of offline handwritten character recognition, they have used gradient direction feature extraction from binary images and from gray-scale images. On the other hand, in online handwriting character recognition, they have extracted the stroke direction features from pen-down trajectory and pen-lifts. In this work, they have performed the classification using the modified quadratic discriminant function (MQDF), nearest prototype classifier (NPC), discriminant feature extraction (DFE) and discriminative learning quadratic discriminant function (DLQDF). The highest test accuracies achieved in their work were 92.1% and 94.9% on the HWDB1.1 (offline) and OLHWDB1.1 (online) datasets, respectively.

Fig. 4 A few samples of online handwritten Chinese characters



2.3 Japanese script

Japanese is an East Asian language spoken by about 126 million people, primarily in Japan country. Japanese has no genetic relationship with Chinese, but it makes extensive use of Chinese characters or kanji in its writing system. Takahashi et al. (1997) proposed a fast HMM based algorithm for online handwritten Japanese Kanji characters. A simple smoothing procedure yields fast and robust learning. After pre-processing steps, the strokes are discretized in a particular manner which naturally leads to a simple procedure for assigning initial state and state transition probabilities. Due to non-iterative learning, the recognition is very fast. They have achieved an average character recognition accuracy of 95.4% for the 881 Kanji characters. Matsumoto et al. (2001) created online handwritten Japanese characters database with the contribution of 163 writers. They have collected 10,000 character patterns, covering 4438 categories in context of sentences. Furthermore, they have analyzed this dataset in term of number of strokes per character, stroke ordering, and their variation. Figure 5 illustrate some samples of handwritten Japanese characters, taken from the dataset.

Jäger et al. (2003) presented state of the art for Japanese online handwritten character recognition with comparison to western handwriting recognition. In their paper, the authors have discussed the crucial developments in pre-processing, classification, and post-processing steps for Japanese character recognition. Western recognizers perform a more complex normalization because of the variable length of western words. They have used the nearest neighbor classifiers for Japanese handwriting and HMMs for western handwriting recognition. Liu and Zhou (2006) proposed efficient trajectory-based normalization and direction feature extraction methods for online handwritten Japanese character recognition. In their work, they have compared one dimensional, pseudo two dimensional normalization methods, and directional features from the original and normalized patterns. To evaluate the performance of the proposed methods, they used TUAT HANDS, kuchibue_d-97-06 and nakayosi_t-98-09 databases of online handwritten Japanese characters. The experimental results show that the pseudo two dimensional normalization methods yield higher



Fig. 5 Some samples of online handwritten Japanese characters dataset

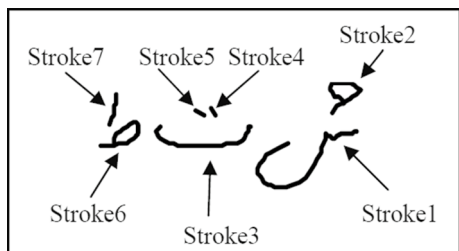
recognition accuracies than one-dimensional methods. Afterwards, Zhou et al. (2007) improved the performance of online handwritten Japanese character string recognition and segmentation accuracies by integrating the geometric context in the path search module. To evaluate the performance of a character string recognition system, they used TUAT HANDS databases. Their experimental work concludes that the geometric features are essential for improving the recognition and segmentation accuracy. The recognition accuracy increased from 87.4 to 94.7% by unary geometry and further increased to 97.1% by binary geometry.

2.4 Persian script

Persian (Farsi) is a member of the Western Iranian branch of the Indo-European language family. Persian script is used by more than 30% of the world's population (Halavati and Shouraki 2007). There are approximately 110 million Persian speakers, mainly in Iran, Afghanistan and Tajikistan, and also in Uzbekistan, Iraq, Russia and Azerbaijan. The official language of Iran is sometimes called Farsi. A little research has been carried out on the recognition of online handwritten Persian script. Persian script is a fully cursive handwriting. Each Persian character may have different forms in different part of word and characters overlapping is more in Persian script. Persian script comprises 32 main characters and written from right to left. The appearance of a character in Persian script is multi-form as begining, middle, end, and isolated form. Persian characters have one main stroke and 1–3 secondary strokes. Usually the main stroke is written before the secondary strokes. In Fig. 6, strokes number 1, 3, and 6 are the main strokes.

Baghshah et al. (2006), the authors proposed a novel approach to recognize the Online Persian handwriting using Fuzzy classifier. The features they have extracted are: Start2End_Direction, Start2COG_Direction, End2COG_Direction, Straightness, Horizontal_Motion, Vertical_Motion, Curvature_Side, and Aspect ratio. They have used fuzzy learning vector quantization (FLVQ) algorithm for stroke classification. Their proposed approach achieved the recognition rate of 95% with tuning the parameters. In Halavati and Shouraki (2007), the online handwritten text is segmented into a sequence of lines, arcs, and half-circles and represent these segments with fuzzy linguistic terms. Features they have extracted are: segment type, segment direction, curvature direction, and segment length. The elastic pattern matching and fuzzy modeling have been implemented in their work to recognize the Persian handwriting. Further, they have compared the performance of their proposed approach with other state-of-the-arts (i.e., Hierarchical Rule Base, Structural Fuzzy, Template Matching, k -NN, and Evolutionary Neuro Fuzzy). Their experiments reported an accuracy rate of 78.0% without using dictionary and 96.0% with using dictionary. Later on, Izadi et al. (2008) proposed a new segmentation algorithm for online handwritten Persian

Fig. 6 An example of main and secondary stroke in Persian script



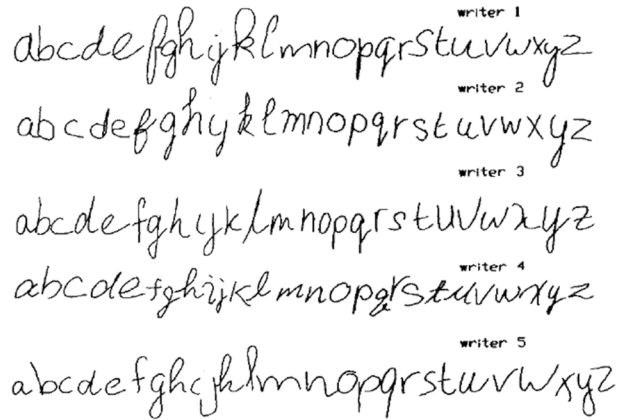
word recognition system. Along with, they presented a perturbation method, which is used to generate artificial samples from the existing online handwritten words. They have used the wavelet-based smoothing technique for pre-processing the data. The word is segmented into convex portion of the global shape, also called Convex Curve Sectors (CCSs). The Dynamic Time Warping based classifier has been employed for Persian word classification in their work. However, there was no publicly available online Persian dataset so, they have used the persian sub-words data in their study (Razavi and Kabir 2004). An average accuracy rate of 89.4% has been achieved for 2-letter word and an accuracy rate of 85.0% achieved for 3-letter word using 1 -NN classifier. Ghods et al. (2013) investigated the effect of delayed strokes (written after the main stroke) on the recognition of online Farsi handwriting. The delayed stroke is studied from two aspects (1) effect on the subword model and lexicon reduction. They have trained the HMM classifier in their experiments. In the experimentation, they have examined 1000 online handwritten Farsi subwords regardless of the delayed strokes and achieved an accuracy rate of 85.2%.

2.5 Roman script

Roman (latin) script is used to write the English language and most widely used and first ranked language across the world. English is an Indo-European language spoken by 1.132 billion speakers across the world. Roman script contains of 26 alphabets and each alphabet has two forms (i.e., upper case and lower case alphabets). The writing structure of Roman script is quite complex. It is written like similar to printed text and cursively. In general, a Roman upper case hand-printed character is written using average about two strokes, lower case character written using one stroke, and in case of cursive handwriting it is written less than a stroke. Tappert et al. (1990) reported a survey on online handwriting recognition for Roman, Chinese, and Japanese scripting languages. In their work, the authors have explored various pre-processing and post-processing techniques used to recognize these scripts. In Roman script there are many characters those are similar in shape and create confusion in classification, such as *U-V*, *C-L*, *a-d*, *n-h*, and *F-E*. Moreover, some characters (i.e., *O*, *I*, *l*, *Z*, and *S*) are also confused with some numbers (i.e., 0, 1, 1, 2, and 5) due to their identical shape. The pre-processing techniques they have discussed in their work are: external segmentation, smoothing, wild point correction, de-hooking, dot reduction, size-normalization, and stroke length normalization. They have extracted shape based features for Roman character recognition. According to them, feature such as ascenders, cusps, closures, etc. and sequences of coded zones can be alphabet's shape specific. The classification methodologies namely decision tree, elastic matching have been used to recognize the Roman characters. Veltman and Prasad (1994), the authors recognized the online handwritten Roman characters using Hidden Markov Models. In their study, the symbols are after pre-processing the input data. The pre-processing steps include, dot detection, dehooking, rotation independency, size normalization, and quantization. To evaluate the performance of their recognizer, they have considered the handwritten dataset, written by five different writers. Each writer asked to write 30 samples for each character in fully unconstrained way. The Fig. 7 illustrating the five samples of different writers.

An average error rate of 6.9% was reported by their recognition system for Individual system and 19.1% for combined system. Moreover, Plamondon and Srihari (2000) investigated various pre-processing and post-processing methodologies used for online and offline handwriting recognition. In pre-processing, the primary approaches they have used to reduce the noise in online handwritten data are: data smoothing, signal filtering,

Fig. 7 An example of main and secondary stroke in Persian script



dehooking and break correction. The character segmentation also discussed in the pre-processing phase. Further, to recognize the characters, they explored structural, rule based, and statistical methods. In their study, they also discussed the n -gram language models in post-processing phase to improve the word recognition rate. Jaeger et al. (2001) presented the online handwriting recognition system, named as NPen++. They have used multi-state time delay neural network (MSTDNN) approach to recognize the online handwritten Roman words. Before extracting the features they have applied several pre-processing steps on the raw data, namely baselines computation, size-normalization, interpolation of missing points, smoothing, slant correction, removal of delayed strokes, and Resampling. The features they have computed, include vertical position of the points, writing direction, curvature, pen-up /pen-down, aspect-ratio, curliness, slope, ascender/descender, and context-map. They have trained the MSTDNN at state level, character-level, and word level. Moreover, they have also reported the efficient tree search and pruning techniques for searching the word from large sized dictionaries. The highest accuracy they have achieved in their experimentation is 96.0% for 5000 words dictionary. Tan et al. (2009) proposed a novel approach for identifying the online handwritten script using information retrieval technique. In their work, they have considered three scripting languages namely, Arabic, Roman, and Tamil and achieved an average recognition accuracy rate of 93.3%.

2.6 Thai script

Thai is the national and official language of Thailand country, natively spoken by over 20 million people. It has 44 Thai alphabets. In the last two decades, a very few amount of research work has been done on the Thai language. Budsayaplakorn et al. (2003) proposed an online handwritten Thai character recognition system. In this system, they have employed two distinct methods: HMM and Fuzzy logic classifier. The construction of fuzzy rule attempted to separate an ambiguous result from HMM classification in their work. To evaluate the performance of the proposed system, they collected 13,608 Thai character samples for training and 7664 character samples for testing. They have shown the improvement in recognition accuracy from 89.0 to 91.2% by incorporating the distinctive feature based Fuzzy classifier with HMM. Later on, Sanguansat et al. (2004) proposed an online Thai handwritten character recognition system using HMM and SVM classifiers. In this work, the authors have experimented SVMs with the score-space kernel to recognize

the online handwritten Thai characters and applied HMM to correctly recognize the confused characters, that were incorrectly recognized by the SVM classifier. Moreover, they proposed the score-space with the symmetric property, called symmetric likelihood ratio score-space, where one observation sequence is mapped to only one score. In this experiment, they have collected 14,557 character samples of 42 Thai alphabets for training the classifier written by 31 writers and 7812 character samples for testing, written by 62 new writers. Their recognition system yields an average accuracy of 96.2% for writer-dependent approach and 92.5% for writer-independent approach by using the combination of HMM and SVM with likelihood ratio score-space. Karnchanapuskij et al. (2009) created an online handwriting Thai character recognition system, which used the linear interpolation approach. This approach helps to design a system to process and analyze Thai handwriting to be converted into textual characters suitable for computer interpretation. They have implemented composition of six features of a Thai character, such as the head of Thai *Consonant*, end of Thai *Consonant*, closed loop, twist, piecewise curve, and straight line. The proposed character recognition algorithm in their work calculates the important information such as angles and rotation directions. XML database is utilized to store the training samples of the character information. In order to test the performance of the algorithm, 80 different Thai characters were used. The recognition results of their system achieved an accuracy of 90.9% for Thai characters.

3 Work done for the recognition of Indic scripts

India is a multi-lingual country, wherein the constitution of India has accepted 22 official languages, namely, Assamese, Bengali, Bodo, Dogri, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Maithili, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Santhali, Sindhi, Tamil, Telugu, and Urdu. Out of which, 10 major scripts (i.e., Devanagari, Bangla, Gurmukhi, Gujrati, Oriya, Kannada, Telugu, Tamil, Malayalam, and Urdu) are used as official documents scripts. These official scripting languages are derived from Brahmi (an ancient script) with various transformations (Ghosh et al. 2010).

Figure 8 illustrates fifteen different Indic-scripts, used for writing these official languages. Most of the Indic-scripts are derived from Brahmi script through various transformations (Datta 1984). A single script can be used to write one or more languages for example, Devnagari is used to write Hindi, Marathi, Nepali, Sanskrit, Konkani, Maithili, Santali, Sindhi, and Kashmiri, while for writing the Assamese and Bengali languages, Bangla script is used. Hindi (spoken by almost 38% of the total population in India) is a widely spoken language and treated as national language of India. Table 2 showing the major scripting languages used in India along with their coverage in various states and number of speakers per language information.

Since, every scripting language has their own structure and has a complex composition of its constituent symbols. Therefore, these scripting languages have their own properties and complexities in handwriting recognition. However, the writing structure of most of the Indic scripts is different from non-Indic scripts as writing structure of Indic scripts' (i.e., Gurmukhi, Devanagari, and Bangla) represented in three horizontal zones (see Fig. 9). It is also noted that many characters of these scripts have horizontal-line, appear above the character. This horizontal-line called as *head-line* in Gurmukhi, *matra* in Bangla, and *shirorekha* in Devanagari. However, the concept of upper- and lower-zone characters is absent in non-Indic script like English. Apart from this there are some more complexities, those

Fig. 8 Examples of 15 Indian official scripts, used for writing

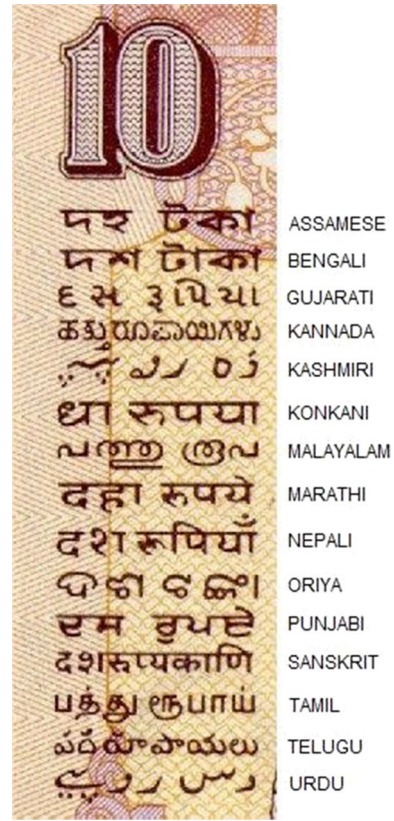


Table 2 Major scripting languages in India with area coverage and total number of speakers per language

Script	Language	Coverage/States	Total speakers (in millions)
Devanagari	Hindi	Haryana, U.P., Rajasthan, H.P., M.P., Bihar	692
Bangla	Bengali	West Bengal, Tripura	107
Telugu	Telugu	Andhra Pradesh	95
Tamil	Tamil	Tamil Nadu	77
Nastaliq	Urdu	U.P., Punjab, West Bangal, Bihar, Karnataka, Maharashtra, A.P. Haryana, Rajasthan	63
Gujarati	Gujarati	Gujarat	60
Kannada	Kannada	Karnataka	59
Oriya	Oriya	Orissa	43
Gurmukhi	Punjabi	Punjab	37
Malayalam	Malayalam	Kerala	36
Assamese	Assamese	Assam	22

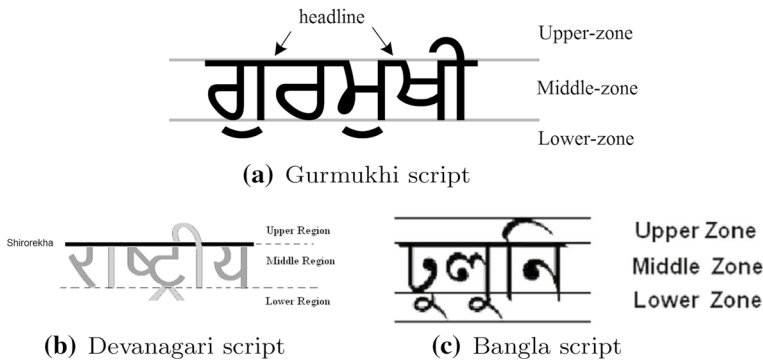


Fig. 9 Zone-wise representation of different scripting languages

are being tackled explicitly in Indian scripts like, stroke/character/matra writing sequence, position of vowels, character rendering order, etc. The next subsections contain the literature review for online handwriting recognition of Indian scripts.

3.1 Assamese script

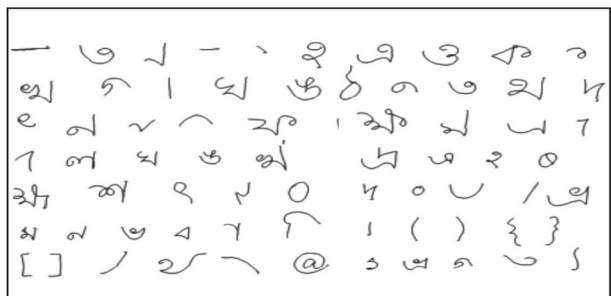
The Assamese script is used to write the Assamese language, and is written from left to right direction. It is the Indo-Aryan language and is mainly used in parts of Arunachal Pradesh and other North-East states of India. It is spoken by 13 million people across the world. The phonetic character set of Assamese has been derived from Sanskrit. Reddy et al. (2012b) have presented a combined online and offline approach for the recognition of Assamese numerals. In their work, the online handwritten numeral recognition system is developed by considering the x -, y -coordinates as the features. On the other side, in offline handwritten numeral recognition system, the considered features included Vertical Projection Profile and Horizontal Projection Profile (VPP-HPP), Discrete Cosine Transform (DCT), Chain Code Histogram (CCH) and Pixel Level features, and Vector Quantization (VQ). The online handwritten captured data is converted into the images for offline system by constructing an image from the x -, y - coordinates. The pre-processing steps they have incorporated in their online handwriting recognition system are: size normalization, smoothing, linear interpolation, and re-sampling. In case of offline recognition system, the input images are converted into binary images. Now, cropping and size normalization is performed where after normalization a 64×64 size image is produced for further processing. The classifiers HMM and VQ are used for classifying the numeral classes for online and offline recognizer, respectively. The average recognition rate of their combined system claimed a significant improvement over the individual online and offline systems. Their combined numerals recognizer produced 99.3% accuracy. Later on, Reddy et al. (2012a) reported the online handwritten digit recognition system using Hidden Markov Models. A large data of 18,000 examples of each numeral have been collected from 100 writers. The development of the handwritten numeral recognition system consists of four stages, viz., pre-processing, feature extraction, modeling, and testing. The pre-processing stage performs size normalization, smoothing, interpolation of missing points, removes duplicate points, and resampling of points. For feature extraction, the pre-processed points (x -, y - coordinates) of a handwritten stroke and the first and second derivatives of x - and

y- coordinates of each point are considered as features. They have used 50% data for training the classifier and the remaining 50% for testing the classifier. Their Assamese handwritten numerals recognition system produced an average recognition performance of 96.0%. Prasanna et al. (2013) developed an Assamese online handwriting recognizer for numerals and isolated aksharas. A list of 240 Assamese aksharas (11 vowels, 40 consonants, 147 conjuncts, 10 numerals, 10 vowel modifiers, 2 consonant modifiers and 20 special symbols) has been considered for the recognition. They have dealt with strokes, sub-strokes and suprastrokes to form the Assamese aksharas. The pre-processing steps: removal of duplicate points, size normalization, smoothing, interpolation of missing points and resampling have been applied on the input data. In feature extraction, the combination of pre-processed x-, y-coordinates, first and second derivatives of x- and y-coordinates for each point are considered as features. The Hidden Markov Model (HMM) classifier has been employed in their work for training the models and recognizing the labels. Their online handwritten isolated Assamese numeral recognizer produced an accuracy of 95.2% and they reported the accuracies of 81.7%, 81.6%, and 85.5% for stroke, sub-stroke, and suprastroke classifiers, respectively. After combining these three classifiers, an average accuracy of 81.8% has also been achieved by them. Mandal et al. (2015) have proposed a curvature point detection based technique to predict variable number of states for modeling a handwritten stroke for online handwritten Assamese script. They have claimed 93.6% accuracy using the proposed curvature point approach. Moreover, Choudhury et al. (2015) investigated a technique to recognize the online handwritten Assamese characters using HMM and SVM stroke classifiers. In their work, they identified 212 isolated stroke shapes. For training and testing the classifier, they have considered 200 samples per stroke class. Figure 10 illustrates some of the examples of handwritten Assamese stroke classes.

3.2 Bangla script

Bangla is an Indo-Aryan language, most widely spoken in West Bengal, an Indian state and also in Bangladesh. It is the second most widely spoken language among the 22 scheduled languages of India. In the recent past, a good amount of research work has been carried out for online handwriting recognition of Bangla script by many researchers in India. Garain et al. (2002) presented a framework of online handwriting recognition for Indian scripts, wherein the authors have experimented with two scripting language, namely, Devnagari and Bangla. They proposed an approach, where the human motor functionality is modeled while writing characters. This functionality is achieved by looking at the whole pen trajectory where the time evaluation of the pen coordinates play a crucial role. The features they

Fig. 10 An example of online handwritten Assamese strokes



have extracted in their work are: angle variation information, euclidean distance between two adjacent points, and 8-directional coding information. Characters are classified by using the template matching approach. Online handwritten dataset of considerable size was used to test the performance of their recognition system. The experiments reported the promising recognition rates such as 97.9% for Devnagari and 96.3% for Bangla script. Bhattacharya et al. (2007) presented a novel direction code based feature extraction approach for the recognition of online Bangla handwritten characters. In their work, they have collected the dataset of 7043 samples of basic online handwritten isolated Bangla characters, written by 114 writers. The pre-processing steps include removing duplicate points and resampling the points. For classification task they have used multilayer perceptron (MLP) (the well known backpropagation algorithm is used to train the MLP classifier) with 70 hidden nodes. The proposed recognition system achieved 93.9% accuracy on training data and 83.6% accuracy on test data. Thereafter, Parui et al. (2008) presented a novel scheme for the recognition of online handwritten basic Bangla characters. A total of 54 distinct stroke classes have been identified for character recognition. The sub-strokes are extracted from a stroke for the feature extraction vector, where 6 scalar features are extracted from each sub-stroke. The HMM classifier has been employed for stroke classification. One HMM is constructed for each stroke class. They have claimed an accuracy of 84.6% for stroke classification. Later on, Bhattacharya et al. (2008) discussed a prototype for online handwritten Bangla cursive word recognition. In this scheme, the online handwritten word is segmented into strokes. A stroke is further divided into 7 sub-strokes of approximately equal length. A histogram of the direction codes is calculated for each sub-stroke as a feature vector. These strokes are further classified using Modified Quadratic Discriminant Function (MQDF) classifier. Their experiments achieved an overall word level recognition accuracy of 82.3%. Bandyopadhyay and Chakraborty (2009) discussed a case study of online handwriting recognition system for Bangla characters. They represent a handwritten character as a combination of strokes. The pre-processing process includes normalization, smoothing and resampling to 50 equidistant points. The features they have included in their work are: 8-directional chain coding and 12 shape features (8 bump points and 4 critical points). Dynamic Time Wrapping based classifier was employed to classify the strokes from the unknown feature strings. The average character recognition accuracy achieved by their simulated experiments is 96.6%. Mondal et al. (2009) attempted an effort

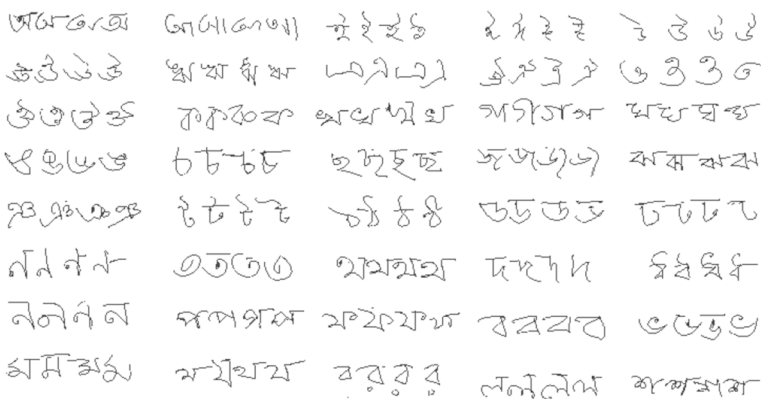


Fig. 11 A few samples of online handwritten Bangla characters

to create the database of online handwritten isolated basic characters for Bangla script. Moreover, they have also discussed the strategy for extracting sub-strokes from the online handwritten character samples, written cursively. They have evaluated their database using two classifiers, namely HMM and Dynamic Time Warping (DTW) based nearest-neighbor classifier. Figure 11 illustrates a few samples of online handwritten Bangla database.

Fink et al. (2010) reported the online Bangla word recognition using sub-stroke level features and Hidden Markov Models. They presented a new approach, which considers the recognition of cursively written words instead of isolated characters. Their proposed scheme achieved a quite promising results for the writer independent online Bangla handwriting recognition task. Mohiuddin et al. (2011) presented an online handwriting recognition system for unconstrained Bangla cursive handwriting, based on combination of Multilayer Perceptron (MLP) and Support Vector Machine (SVM). The MLP architecture is used for sub-stroke level feature extraction and reduction in dimension of the input feature vector; and SVM is used for final recognition. In their study, feature selection process is based on segmentation of the input handwritten word into a fixed number of sub-strokes. The simulation results show that the combination of MLP and SVM improved the recognition accuracy. They have achieved 88.7% recognition accuracy on a test data, containing 50 city names and 87.2% on the test data, containing 110 city names. Biswas et al. (2012) reported the HMM based online handwritten Bangla character recognition using dirichlet distributions. The authors have collected a reasonably large database of online handwritten Bangla characters at stroke level. In their work, they have identified 75 stroke classes. In pre-processing, size normalization and local noise removing have been applied. The two-stage approach is used for character recognition. First, a probability distribution is estimated for each stroke class, and then HMM is applied. Their proposed approach obtained the recognition accuracy of 91.9%. Chowdhury et al. (2013) proposed a novel approach (distance function based on Levenshtein Distance Metric) for online handwritten character recognition. Directional and positional information based features have been extracted for the character recognition in their work. They have considered five databases of different Indian scripts (Bangla-numerals, Bangla-basic characters, Devanagari, Telugu, and Tamil) for testing the performance of their proposed system. The experimental results of handwritten Bangla numerals show an improvement when compared to the existing recognition results. Their proposed recognition system achieved an accuracy of 98.4% for Bangla numerals. Later on, Samanta et al. (2014) proposed a novel approach for online unconstrained handwritten Bangla words recognition using Hidden Markov Model. In their work, a whole word sample is considered as a basic unit for the recognition, instead of recognizing individual sub-strokes. Circular feature, linear feature and combination of both the features are computed at sub-strokes level after segmenting the word into sub-strokes. They have implemented a fully connected non-homogeneous HMMs in their experiments. The smoothing of probability estimates at two levels employed by them resulted into an accuracy of 89.7% by using the combined features with 2L-S approach for online handwritten Bangla word recognition. Bhattacharya et al. (2017) addressed an approach to detect and clean the various types of noises in online handwritten captured data. They have reported the drastically improvement in the recognition accuracy using aforementioned approach. Later on, Bhattacharya et al. (2018a), reported a writer-independent online handwritten Bangla word (cursively written) recognition system addition with modified features, named as Mod-NPen features. The Mod-NPen feature set includes direction, curvature, aspect ratio, curliness, slope and context maps features. They have evaluated their work by considering the variable size of vocabulary (i.e., 500–20,000 word lexicons). The highest accuracy rate of 93.4% has been obtained in their experiments on a test set of 250 different word classes.

Additionally, in (Bhattacharya et al. 2018b), the authors have proposed a novel *sub-stroke-wise feature* (SRF) based approach to recognize the online cursive handwritten Bangla and Devanagari words. Their proposed recognition system significantly outperforms the existing feature sets for online handwritten Bangla and Devanagari words. In their experimentation, an accuracy rate of 94.1% achieved for Bangla words and 88.1% for Devanagari words, written cursively.

3.3 Devanagari script

Devanagari script is widely used in India and Nepal and is written from left to right direction. In India, the Devanagari script is used for writing several major languages such as Sanskrit, Hindi, and Marathi. In Nepal, Devanagari script is used to write the Nepali language. Connell et al. (2000) presented a detailed study on the recognition of unconstrained online handwritten Devanagari characters. They have employed five different classifiers for the recognition task. To test the performance of their proposed system, they have considered a total of 1600 Devanagari characters from 40 classes, written by 20 writers. The combined classifier achieved an average character recognition accuracy of 86.5% with no rejects. Garain et al. (2002) presented an online handwriting recognition system for Indian scripts, wherein they have experimented with two scripting language, namely, Devanagari and Bangla. In their work, the human motor functionality is modeled while writing characters. The features they have extracted in their work are: angle variation information, euclidean distance between two adjacent points, and 8-directional coding information. Characters are classified using template matching approach. Online handwritten dataset of considerable size was used to test the performance of their recognition system. The experiments reported an average recognition accuracy of 97.9% for Devanagari characters. Namboodiri and Jain (2004) proposed an algorithm to classify words and lines from an online handwritten document into one of the six majorly used scripts: Arabic, Devanagari, Cyrillic, Han, Hebrew, and Roman. In feature extraction, 11 different spatial and temporal features have been extracted from the strokes of handwritten words. They have employed K -NN, neural net, and SVM-based classifiers in their work. Their proposed recognition system attained an overall classification accuracy of 87.1% at word level with fivefold cross-validation on a dataset of 13,379 words. Thereafter, Joshi et al. (2005) presented a writer-dependent system for automatic recognition of isolated handwritten Devanagari characters. Syntactic and structural feature-based approaches have been used to recognize the handwritten characters. A dataset, containing 1487 Devanagari characters have been used to test the performance of the recognition system. The experimental results showed a recognition accuracy of 88.9% for the test data, wherein the shirorekha was removed and an accuracy of 87.1% for the test data without removing the shirorekha. Swethalakshmi et al. (2006) proposed an online handwritten character recognition system for Devanagari and Telugu scripts. The three different pre-processing steps, namely, normalization, smoothing, and interpolation have been used to pre-process the handwritten strokes data. Further, these strokes are classified using the Support Vector Machine (SVM) classifier, where one-vs-many multiclass classification strategy is implemented for stroke classification. In their work, a total of 91 stroke classes have been identified and considered for classification. In the post-processing phase, after the stroke recognition, a rule-based approach is used to form a Devanagari character from the recognized strokes. The stroke classification results show that with 46 stroke-classes and 60 features, their recognition system has achieved 96.7% accuracy and on the other hand with 82 stroke-classes and 120 features the

recognition system attained 97.3% accuracy. Mondal et al. (2010) reported the online handwritten isolated character recognition system by using the existing benchmark of four major scripting languages of India, namely, Bangla, Devanagari, Tamil, and Telugu. These standard benchmark databases are freely available for research competitions. They have considered a set of 111 devanagari character-symbols, including, basic characters, character modifiers, frequently occurring conjuncts and half-form consonants. The Devanagari character database, consisting of 23,891 characters, written by 109 different writers has been used for training and testing the recognition system. The features extracted in their work are: (1) chain code histogram feature, and (2) point-float feature. For character classification, they have used the Nearest Neighbour (NN), Multi-Layer Perceptron (MLP), and Hidden Markov Model (HMM) classifiers. The highest Devanagari character recognition accuracy of 95.3% has been achieved with Nearest Neighbour classifier. Belhe et al. (2010) built a semi-automatic annotation tool for annotating online handwritten data of Indic scripts. They have proposed a XML standard for representing the online handwritten data, wherein they have described the annotation at stroke-, character- and word-level. A few samples of their Devanagari character database are shown in Fig. 12. They have used this tool extensively for annotating a large amount of data of Devanagari script. Kumar and Bhattacharya (2010) presented a novel scheme for the recognition of online handwritten basic isolated characters for Devanagari script by using the Hidden Markov Models. They built a HMM-based stroke classifier for 42 stroke-classes. The pre-processing steps applied to the sequence of handwritten strokes were smoothing and interpolation. In their work, a stroke is segmented into several sub-strokes by categorizing the stroke s_i as N (North), or S (South), or E (East), or W (West) labels. Further, they extracted equidistance points from the sub-stroke by using the scalar feature extraction approach. The look up table approach is employed to classify the final character from the recognized strokes. Bharath and Madhavanath (2012) proposed HMM-based lexicon-driven and lexicon-free word recognition for online handwritten Devanagari and Tamil scripts. In their work, they have discussed various stages of developing the recognition system such as symbol set definition and data set creation, pre-processing steps, feature extraction techniques, and HMM based modelling of strokes. Moreover, they have also discussed the lexicon-driven and lexicon-free strategies for the Devanagari word recognition. According to the standard writing order, a symbol tree is modeled and the Viterbi decoding strategy is used to recognize the input word. On the other hand, a recurrent HMM and a novel “Bag-of-Symbols” representation and matching scheme is used in the lexicon-free strategy to recognize the input words. The experimental result achieved the accuracies 93.4% and 87.1% for 1000 and 20,000 lexicons, respectively for Devanagari script with the combined lexicon-driven and lexicon-free strategy. Belhe et al. (2012) have proposed HMM and symbol tree based online handwritten isolated Hindi words recognition system. In their system, they convert the online stroke

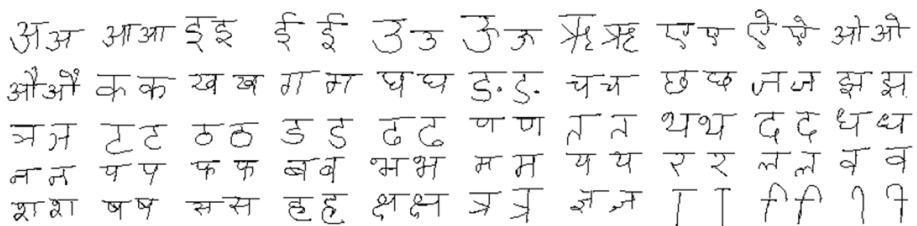


Fig. 12 Samples of online handwritten Devanagari characters

information into an offline image during the pre-processing phase. These offline images are further processed for feature extraction. Histogram of Oriented Gradients (HOG) feature vector is extracted for each image. They have reported an accuracy of 89.0% in their work. Mehrotra et al. (2013) presented a novel offline strategy for recognition of unconstrained online handwritten Devanagari characters. A list of 460 Devanagari words has been considered for data collection, which covers all the possible combinations of aksharas/syllables in Devanagari script. In their work, the collected data have been annotated at word-, akshara-, and stroke-level. The pre-processing steps followed in their work are: character segmentation, normalization, up sampling, and dilation. The Convolutional Neural Networks (CNNs) classifier has been employed for character classification. They have experimented with 10 different configurations of CNN model in order to build a better training model. They have claimed an average character recognition accuracy of 98.2% on test data. Moreover, in Ghosh et al. (2019) the authors reported a novel Dempster–Shafer theory (DST) based biometric approach for online handwritten signature recognition and verification system for Devanagari and Latin script. Their signature recognition system classify the handwritten samples using Hidden Markov Model (HMM) and Support Vector Machine (SVM). In experimentation, the signature recognition results outperform other state-of-the-art.

3.4 Gurmukhi script

Gurmukhi script is used for writing the Punjabi language. Punjabi is an Indo-Aryan language spoken by 102 million speakers across the world. Gurmukhi script has 35 basic *Consonants*, 6 modified *Consonants*, and 10 *Vowels* (Bahri 1982). As per our knowledge, a very less amount of research work has been carried out on online handwriting recognition of Gurmukhi script in last two decades.

Initially, the research work on online handwritten Gurmukhi script was started with the basic Gurmukhi characters recognition using the elastic matching scheme (Sharma et al. 2008). In their work, the authors have described the process of Gurmukhi character recognition in two stages. In the first stage, process of the online handwritten strokes recognition has been discussed and afterwards the process of formation of the Gurmukhi character from the recognized strokes has been discussed. The online handwritten data samples have been collected for 41 Gurmukhi characters. A total of 50 stroke classes were identified to recognize these 41 Gurmukhi characters. The XML file format is used to store the strokes information (x -, y -coordinate points) associated with a character. Their proposed system achieved an average recognition accuracy of 90.1%. In an another attempt, Sharma et al. (2009a) explored the online handwritten Gurmukhi strokes pre-processing algorithms for identifying the improvements in the recognition of four high-level features (loop, headline, straight line, and dot) of Gurmukhi strokes. In their work, the authors implemented the pre-processing algorithms: size normalization and centering; interpolating missing points; smoothing; slant correction; and resampling of points. After incorporating the pre-processing algorithms, the feature-wise stroke recognition accuracies have been improved by 5.0%, 3.3%, 6.7%, and 8.3% for loop, headline, straight line, and dot features, respectively. Later on, Sharma et al. (2009c) presented an another approach to recognize the online handwritten Gurmukhi characters, wherein the character recognition is carried out using the combination of small line segments, chain code, and elastic matching techniques. They reported an improvement of 4.5% in the recognition accuracy on the same dataset of 2460 Gurmukhi characters, written by 60 different writers. Thereafter, Sharma et al. (2009b)

extended their work by proposing the online handwritten Gurmukhi word recognition by rearranging the recognized stroke. In the rearrangement process, each recognized stroke is identified as dependent or major dependent stroke for a character and the respective position of the stroke on x - and y -axis is computed. The pre-processing steps include, size normalization and centering, interpolating missing points, smoothing, slant correction, and resampling of points. In this work, high level features (i.e., loop, crossing, straight line, headline, and dots) are computed on the basis of low level features. They have used elastic matching technique to recognize the online handwritten strokes. To test the performance of their proposed system, a dataset of 2576 Gurmukhi words, written by 11 writers have been used. Among these words, 2087 words were recognized correctly with an overall recognition rate of 81.0%. Kumar and Sharma (2013) presented an efficient post-processing algorithm for online handwritten Gurmukhi character recognition, wherein the Gurmukhi characters are formed by using a Rule-based approach. They have identified a total of 114 stroke classes for stroke classification. For testing the performance of their proposed system, they have considered a dataset, consisting of 184 samples of each 45 Gurmukhi characters. The proposed algorithm achieved the promising recognition accuracy of 95.6% for single character stroke sequence. Thereafter, Kumar et al. (2015) proposed an algorithm for online handwritten Gurmukhi akshara formation from the recognized strokes by using the Support Vector Machine (SVM) classifier. The pre-processing steps include, size-normalization, de-hooking, smoothing, and resampling. They have considered the resampled 64 points (x -, y -coordinates) as the feature vector of size 128. This feature vector is further normalized in the range [1,9] using SVM-scale function. A total number of 114 stroke classes have been identified for stroke classification. For training the classifier they have collected the akshara based handwritten data from 148 different individuals. To test the performance of their post-processing akshara formation algorithm, they used a test data set of 4310 Gurmukhi aksharas, written by ten different writers. Their proposed system achieved an overall recognition accuracy of 80.4% for Gurmukhi aksharas. Singh and Sachan (2015) reported a framework of online handwritten Gurmukhi script recognition. The pre-processing steps included in their work are: size normalization and centering; identification of missing points; stroke smoothing; and resampling of points. They have discussed two different classification techniques: (1) structural and Rule-based methods, and (2) statistical classification methods. The nearest neighbour and SVM classifiers have been employed for the stroke classification and they achieved 86.9% accuracy at Gurmukhi word level. Verma and Sharma (2015) presented an analysis on the performance of various zone based features for online handwritten Gurmukhi script. In their work, five zone based features: normalized features, diagonal features, directional features, parabola based curve fitting features, and power curve based features have been experimented. They have characterized the handwritten Gurmukhi script in three horizontal zones, namely, upper, middle, and lower zones. They have identified 12 stroke-classes for upper zone, 82 for middle zone, and 7 for lower zone. For training the classifier, they have considered 100 samples per class. They performed the experiments with all the standard kernels of SVM classifier (linear, polynomial, radial basis function, and sigmoid) with k -fold cross-validation scheme. Their experiments yielded 92.1% accuracy with fivefold cross validation for the classification of middle zone strokes. Afterwards, Verma and Sharma (2016) reported a voting-based online handwritten character recognition system for Gurmukhi script wherein HMM- and SVM-based stroke classification was carried out. The authors identified 74 stroke-classes for Gurmukhi script recognition. The pre-processing steps employed in their work include noise removal, normalization, missing point interpolation and re-sampling. They have experimented with five different features: (1) normalized x -, y -traces, (2) region-based

features, (3) curvature features, (4) curvature feature-based classes, and (5) direction features. They have built a single classifier to classify the strokes in three different zones, namely upper, middle and lower zones. Further, they have tested a data set of 1750 Gurmukhi character samples in order to validate the performance of their recognition system. The authors claimed an accuracy of 96.7% on 35 Gurmukhi characters in the implementation of their recognition system. In another attempt, Verma and Sharma (2017b) have proposed an algorithm, where a Gurmukhi character is recognized using three different zone-wise classifiers. They have considered a total of 99 stroke-classes for three zones (12 for upper zone, 80 for middle zone, and 7 for lower zone). A zone identification algorithm has been proposed in their work, which decides the zone of an input stroke. The strokes are classified into three horizontal zones by using this zone identification algorithms. Further, the grouped strokes in these separate zones are recognized using the respective zone-wise classifier. A Rule-based post-processing approach is used to form a Gurmukhi character from the recognized strokes. The authors claimed an average of accuracy 74.8% for Gurmukhi character recognition. Singh et al. (2016) introduced a novel technique to generate handwritten stroke classes based on limited set of Gurmukhi words. In their study, the online handwritten Gurmukhi word is recognized by grouping the handwritten strokes. For the development of handwritten data set, a total of 33 common names of places of Punjab state have been used. A minimal data set of 39,411 strokes have been collected for 72 stroke classes from handwritten words. The verification of expert-writer or moderate writer has been done using *k*-means clustering technique. Their proposed system achieved recognition results using the Hidden Markov Models (HMMs) as 87.1%, 85.4%, and 84.3% for the middle zone strokes when used the training data as 66.0%, 50.0%, and 80.0% of the developed handwritten dataset.

Singh et al. (2018a) reported the recognition of online handwritten Gurmukhi characters in two horizontal zones instead of three zones (Kumar and Sharma 2013; Kumar et al. 2015; Verma and Sharma 2015, 2017b). A total of 93 stroke-classes (12 for upper zone and 81 for lower zone) have been identified for stroke classification in both the zones. We have experimented three features, namely, pre-processed *x*-, *y*-coordinates, Discrete Fourier Transform, and Directional features. A data set of 52,500 word samples have been collected by 175 writers for training the classifier. The proposed zone identification algorithm achieved an accuracy of 99.8% to correctly classify the strokes in two zones. An attempt to provide the Since, standard dataset of online handwritten Gurmukhi script was not available publically. Therefore, Singh et al. (2019) created a common platform and made the benchmark dataset publically available for the researchers. Figure 13 illustrates five different words samples, taken from the dataset. These words are written by five different writers.

3.5 Kannada

The Kannada script is primarily used to write the Kannada language. It is one of the Dravidian languages of south India. Kannada script is derived from Sanskrit and it is most widely spoken in the state of Karnataka. The Kannada script consists of 16 vowels and 36 consonants. The writing style of the Kannada script is left to right. Prasad et al. (2009) proposed an online handwritten Kannada character recognition based on divide and conquer strategy. The main aim of employing this strategy was to reduce the number of character combination classes. One or more (maximum three) consonants can combine with a vowel to produce a new grapheme. Therefore, a typical Kannada character can be a vowel (V),

Gurmukhi Word	Writer 1	Writer 2	Writer 3	Writer 4	Writer 5
ਉਤੇ					
ਸਾਫ					
ਕੁੱਝ					
ਖੰਡ					
ਉਪਰ					

Fig. 13 Examples of online handwritten words samples, written by five different writers

or a consonant (C), or a CV combination, or a CCV combination, or a CCCV combination or a numeral. So, there are a total of 6,47,921 combinations of Kannada character and it is almost impractical to train a classifier for such a huge number of classes. In their work, a Kannada character is characterized vertically into three regions, namely, middle, top and the bottom. The pre-processing involves noise removal, re-sampling, and size normalization. After segmenting the strokes into three defined regions, the extracted features such as Normalized Horizontal and Vertical Coordinates (a_i and b_i), Normalized Trajectory Features (r_i and θ_i), and Normalized Deviation Features (a_{di} and b_{di}) are mapped to sub-space using PCA. They employed k -NN classifier to classify these three regions. The performance of online handwritten kannada character recognition system for 283 classes was fairly good, with a maximum recognition accuracy of 81.0%. Nethravathi et al. (2010) created a huge (1,00,000 words) annotated dataset for Kannada and Tamil online handwriting recognition systems. They have collected the dataset with the help of 600 writers in order to cover all the possible variation in writing style. They described that Kannada characters are written separately without much overlapping between them and the modifiers are written below or above or adjacent to the base character. They have created a reduced symbol list which includes all the basic symbols of the character and finally, they have considered 295 Kannada symbols for training the classifier. They have used Tablet-PC for the online handwritten data collection. The collected data is further stored and annotated in a standard XML format proposed by OHWR consortium (Belhe et al. 2009). They have also described a semi-automated annotation tool, which helps in annotating the handwritten strokes data to the word level. Kunwar et al. (2010) proposed a novel heuristic approach, (1) to segment the recognizable symbols from the online handwritten Kannada word and (2) to perform recognition of the complete word. In feature extraction, some high level structural features were extracted from the pre-processed stroke group data. They have computed two estimates from the first derivative of each point. They have employed Statistical Dynamic Space Warping (SDSW) classifier to recognize the Unicode character and achieved an average accuracy of 80.0% at Unicode level for Kannada script. Murthy and Ramakrishnan (2011) proposed a novel technique for online handwritten Kannada characters recognition wherein the best classifier is chosen from set of three classifiers in order

to make the recognition system fast and efficient. The pre-processing involves size normalization, smoothing, and re-sampling the points. In their work, they have computed the features, namely, pre-processed x -, y -coordinates, quantized slope between two consecutive points, dominant points, and quartile features. The prototype-based classifier, Dynamic Time warping (DTW), is used for stroke classification. Further, for dimensionality reduction the Principal Component Analysis (PCA) is used to map the original N dimensional feature space to M dimensional space such that $M < N$. They built the classifier for 295 classes and achieved an average accuracy of 77.2% by using a single stage classifier and an average accuracy of 92.7% with the dexterous classifier. Rampalli and Ramakrishnan (2011) presented an online handwritten character recognition system for Kannada script where the online handwritten information is converted into the offline images. The recognition is carried out for both online and offline systems. The handwritten data has been collected from 69 different writers. The data is captured using the Tablet-PC and further pre-processed by applying noise removal, size normalization and resampling the points. The features derived from the online handwritten samples are: x -, y -coordinates, pen direction angle, and first and second derivatives of x - and y -coordinates. However, for offline recognition four different features vectors, namely, Directional Distance Distribution (DDD), distance of Nearest Stroke Pixels (NSP), Transition Count (TC), and Projection Profiles (PP) are computed. They have trained two different SVM-based classifier for the construction of online and offline recognition. Further, they combined the online and offline classifiers output in order to improve the recognition accuracy. They performed the experiments on two different sets (200 and 295 classes) of Kannada character classes. The performance of their fusion classifier produced the recognition accuracy of 92.3%.

Thereafter, Ramakrishnan and Shashidhar (2013) reported the online handwriting recognition system for unrestricted kannada words as most of the existing systems have been reported the problem of recognizing isolated characters. In their work, they have used Attention Feed-based Segmentation (AFS) method to segment the online handwritten Kannada words into its constituent symbols. They have considered a total of 8,63,848 possible

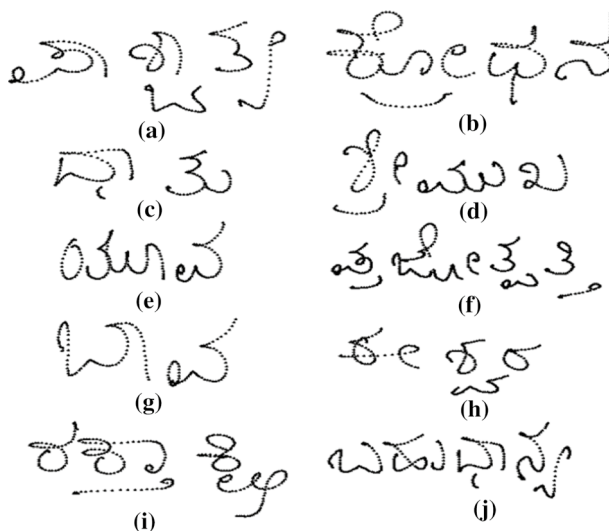


Fig. 14 An example of online handwritten Kannada words

combinations of Kannada consonants and vowels for training the classifier. They have used the same handwritten data set for 295 symbols, collected from 69 writers. The pre-processing including, smoothing, normalization and re-sample have been applied in their work. Figure 14 illustrates some samples of handwritten word of Kannada script. The stroke recognition is performed using Support Vector Machine (SVM) classifier. Their proposed system segmented 42085 words correctly. An average accuracy of 94.3%, 62.0%, and 18.2% achieved for word segmentation, symbol recognition and word recognition, respectively.

3.6 Malayalam script

Malayalam script is a Brahmic script and is commonly used to write the Malayalam language. It is closely related to the Tamil and Sanskrit languages and the principal language of Kerala state of India. This language is spoken by 35 million people across the world. Malayalam script is also used to write the Sanskrit texts in Kerala. Malayalam script consists of 13 vowels, 36 consonants and 5 pure consonants. In addition to them, three other symbols viz., Anuswaram, Visargam and Chandrakkala are part of the Malayalam script. Gowri Shankar and Chakravarthy (2003) presented an online handwritten recognition system for Malayalam characters. They have extracted shape features of handwritten strokes for training the model. A set of 18 shape features have been used in their work. The string matching technique has been employed in order to classify a handwritten stroke. Thereafter, a character is formed by grouping the recognized stroke(s). They tested the performance of their recognition system by considering a data set of 86 strokes with a total of 216 variations. Their experimental results achieved an average stroke recognition accuracy of 90.8%. Figure 15 illustrates a sample of Malayalam handwritten text.

Arora and Namboodiri (2010) reported a hybrid model for the recognition of online handwritten characters for Malayalam and Telugu scripts. In their work, they have discussed the complete process of word formation from the basic building block in an online handwriting recognition system (i.e., a stroke). The pre-processing steps, namely, removing noise, size normalization, smoothing, and resampling of points have been incorporated to remove the variability in the stroke shapes. The features they have extracted in their work are: raw x - and y -coordinates of resampled points, strokes movement upto 4th order, direction and curvature of stroke, length of stroke, aspect ratio, area of stroke, number and direction of points in different windows, projection histogram, and fourier coefficients of x and y sequences. The handwritten strokes have been classified using Hidden Markov Models (HMMs) and top-N probable classes were computed for each stroke along with their probabilities. To compute the most likely sequence of stroke labels, Viterbi-decoding process has been employed. A data set consisting of 120 Malayalam words has been considered to test the model. A data set of 7348 samples of strokes have been used to train the model for 90 Malayalam stroke classes. An overall recognition accuracy of 78.1% has been

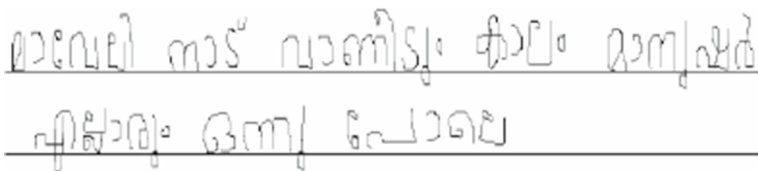


Fig. 15 An examples of online handwritten Malayalam text

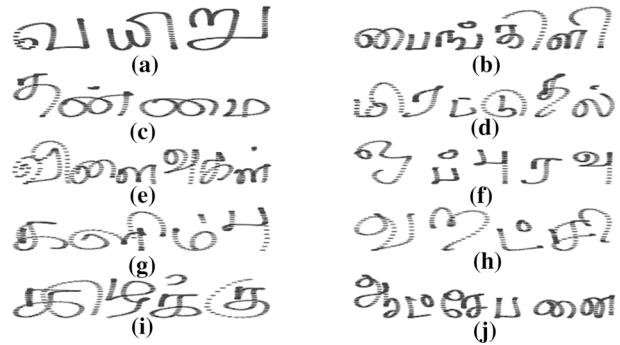
achieved for Malayalam script, when tested on a data set of 60,492 words, collected from 367 different writers. Primekumar and Idiculla (2011) proposed a online handwritten character recognition system using wavelet transform and Simplified Fuzzy ARTMAP (SFAM) for the Malayalam script. The pre-processing steps incorporated in their work are: elimination of duplicate points, smoothing, size normalization and equidistance re-sampling of points. In the feature extraction, first, sequence of relevant features have been extracted from the normalized x -, y -coordinates. Thereafter, the wavelet transform of these features is calculated to form a resulted feature vector in the compressed form. A self organized neural network, SFAM is used for training the model. They built a training model by using 2530 samples for 63 Malayalam symbol classes. This training data set was collected from 40 different individuals. For testing, a new data set consisting of 1279 symbol samples was collected from 20 different individuals. The performance of the system was analyzed using “db1” and “haar” wavelets. The authors claimed a maximum character recognition accuracy of 97.8% by using “haar” wavelete. Later on Indhu and Bhadrans (2012) used Simplified Fuzzy ARTMAP artificial nerual network approach to recognize the online handwritten Malayalam characters. The word-level handwritten data from 29 different writers has been collected to test the performance of proposed recognition system. The structural and directional information from the online handwritten stroke have been extracted for the feature vector. Their proposed writer independent system has achieved an average stroke recognition accuracy of 98.3%. Namboodiri and Team (2013) presented the development process of online handwriting recognition system for Malayalam script. Here, they have collected over 1,00,000 Malayalam words handwritten data from 700 writers. They have considered a total of 130 classes for training the model. In this work, three new feature extraction methods: (1) inverse curvature re-sampling, (2) circle based representation of strokes, and (3) histogram of oriented gradients have been developed. They have used SVM-DDAG classifier to build the training model for character recognition. In addition, they have also used a GPU based implementation of LeNet-5 Convolutional Neural Network. In the post-processing phase, they have used dictionary based lookup tables with a limited dictionary and bigram language model approaches. Their experiments achieved the maximum stroke recognition accuracy of 96.3% when equidistant sampling combined with curvature weighted sampling is used.

3.7 Tamil script

Tamil is the most widely spoken language of south India, Sri Lanka, Malaysia, and Singapore. Tamil script is used to write Tamil language. Tamil script has 12 vowels and 23 consonants. It is written from left to right. Tamil is relatively simpler than other Indian scripts due to having less character compositions. Aparna et al. (2004) presented an online handwriting recognition system for Tamil characters. The captured handwritten data is processed through several pre-processing steps: size normalization, smoothing, and interpolation in their work. They have extracted a total of 18 shape features from the strokes data which include dot, line terminals, bumps, and cusp etc. They have used a soft-matching approach for the stroke identification. A Tamil character is formed by grouping predicted stroke labels. In order to test the performance of their recognition system, a data set 2000 stroke samples, written by 15 different writers is used. They have built a classifier, consisting of 96 stroke classes. Their experimental results achieved an average stroke recognition accuracy of 82.8% for 96 stroke classes. Deepu et al. (2004) presented the Principal Component Analysis (PCA) for online handwritten isolated Tamil charcter recognition. Initially,

the information captured online is processed for smoothing and size normalization. Thereafter, a fixed size feature vector is constructed. The pattern classification is then performed to classify the character class wherein PCA-based classification, PCA with pre-clustering, and PCA with modified distance measure are evaluated. The PCA-based classification schemes are computed and compared with Nearest Neighbor (NN) classifier in their work. Their proposed character recognition system achieved an average accuracy of 90.8% using PCA with modified distance measure scheme. Joshi et al. (2004) presented the comparison of elastic matching algorithms for online handwritten isolated Tamil characters. A data set of 31,200 isolated character samples is collected from 20 writers. Further, this data set is pre-processed through size normalization, smoothing and resampled to 60 points. This pre-processed data is used for training the classifier. They have performed experiments on three different features, namely, pre-processed x -, y -coordinates, quantized slopes, and coordinates of dominant points. They have performed seven schemes to test the performance of their system. Among these schemes, dominant points based two-stage scheme and combination of rigid and elastic matching schemes give a good performance with an average recognition accuracy of 94.8% and 95.9%, respectively. Bharath and Madhvanath (2007) proposed an online handwritten Tamil word recognition system using Hidden Markov Models. They have identified a total of 84 symbol classes for Tamil word recognition. The pre-processing stages involved in their work are: removal of duplicate points, smoothing, and size normalization. They have computed the angle features to capture the writing direction and curvature of the trajectory points. A list of 80 Tamil words has been selected for collecting the handwritten data samples. A data set of 7233 word samples were collected, written by 132 writers. The collected handwritten data is stored in the UNIPEN format (Guyon et al. 1994). Among this collected data, 6,252 word samples were considered for training the model and 981 word samples were considered for testing. Their experimental results achieved an average accuracy of 98.0% for a lexicon size of 1000 and 92.2% for a lexicon size of 20,000. Sundaram and Ramakrishnan (2008) presented the online handwritten Tamil character recognition system using a novel 2-Dimensional Principal Component Analysis (2DPCA) approach. In their work, a novel set of features, namely, radial distance and polar angle, radial distance from quartile mean and polynomial fit have been extracted from the conventional normalized x -, y -coordinates for each character sample. A data set of 23,400 character samples has been collected by them from 15 native Tamil writers. Their experimental results indicate that the proposed 2DPCA approach exhibit 3.0% improvement over the conventional PCA technique. Sundaram and Ramakrishnan (2009) proposed a script specific post-processing approach in order to improve the handwritten character recognition for Tamil script. In their work, they have used the 2DPCA approach with the Neural Network (NN) classifier. The structural cues scheme has been employed to identify the confusion between character classes. Their efforts resulted into a significant improvement in the classification accuracy. They have achieved an average character recognition accuracy of 86.5% on the IWFHR test data set. Thereafter, Sundaram and Ramakrishnan (2013) proposed a lexicon-free, script-dependent approach to segment the online handwritten isolated Tamil words into its constituent symbols. Some online handwritten words samples, taken from MILE word database are depicted in Fig. 16. In their work, the online handwritten word is first segmented into stroke groups using the Attention Feedback Segmentation (AFS) and Dominant Overlap Criterion Segmentation (DOCS) strategies. The classification task is carried out using the SVM classifier. To test the performance of their proposed approach, they have collected a data set of 10,000 isolated handwritten words, containing 53,246 Tamil symbols. They have achieved the symbol-level segmentation accuracy of 98.1%, which has been improved to 99.7% after incorporating the AFS

Fig. 16 Online handwritten Tamil words samples, taken from MILE word database



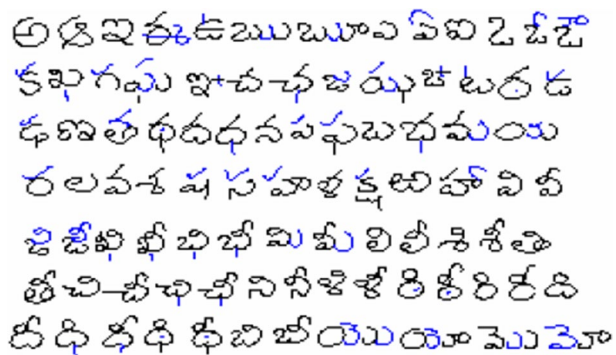
strategy. Additionally, they achieved the symbol recognition accuracy of 83.9% by using DOCS module and of 88.4% after incorporating the AFS module. The word recognition accuracy achieved by DOCS and AFS modules are 50.9% and 64.9%, respectively. Ramkrishnan and Urala (2013) discussed the efficacy of using local features (pre-processed x , y -coordinates), global features (DCT, DFT), and combination of both the local and global features for the recognition of online handwritten Tamil numerals and characters. The SVM classifier based model has been trained using the IWFHR 2006 Tamil handwritten character recognition dataset. Their combined features approach attained an average character recognition accuracy of 95.9% for Tamil script. Later on, Sundaram and Ramakrishnan (2015) presented a post-processing strategy for online handwritten isolated Tamil words. In their work, the handwritten input word is first segmented into its individual symbols and these symbols are recognized using the SVM classifier. Afterwards, they have incorporated the bigram language model at symbol and character level to improve the recognition accuracy. In addition, the expert classifiers have also been employed for reevaluating and disambiguating the different set of confused symbols. The Dynamic Time Warping (DTW) approach has been used for disambiguating the confused symbols. A data set of 15,000 handwritten isolated Tamil words has been considered to test the performance of the proposed bigram language model. Their experimental results show the recognition accuracies of 93.0% at symbol-level and 81.6% at word-level.

3.8 Telugu script

Telugu script is used to write the Telugu language, a Dravidian language and spoken in the Indian states of Andhra Pradesh and Telangana along with some neighbouring states. The Telugu script is also widely used for writing Sanskrit text. Telugu script has 35 consonants, 18 vowels and 3 diacritics. Swethalakshmi et al. (2006) proposed an online handwritten character recognition system for Devanagari and Telugu scripts. In their system, the handwritten strokes are classified using SVM classifier with one-vs-many multiclass classification approach. They have collected the online handwritten data for 253 symbol classes from 92 different writers. Among those writers data, 82 writers data (33726 samples) is used for training and remaining 10 writers data (4091 samples) is used for testing. They trained the SVM classifier by using the Gaussian kernel, which exhibited a better performance. The experimental results show that the SVM classifier yields an average stroke recognition accuracy of 83.1% with the parameters $\sigma = 30$, and $C = 50$ for Telugu script. Babu et al. (2007) presented an online handwritten recognition system for Telugu script. A

total of 141 symbols were identified for the recognition that cover the Telugu script character set. The data set of 38,393 symbol samples were collected from 143 different individuals, belonging to different age groups, genders, and educational backgrounds. The pre-processing steps involved in their work are: removal of duplicate points, smoothing, size normalization, and resampling of points. After pre-processing, a 19 dimensional feature vector (11 time-domain features and 8 frequency domain features) is computed for each point in a sample. In time-domain, the computed features include normalized x -, y -coordinates, normalized first derivatives, normalized second derivatives, curvature, aspect ratio, curliness, and liness. The Discrete Cosine Transform (DCT) feature vector is calculated to determine the frequency-domain features. They have employed the HMM classifier for symbol classification in their work. From the collected data set, they considered 29,158 samples written by 108 writers for training and 9235 samples written by 35 different writers, considered for testing. Their proposed online handwritten recognition system for Telugu script achieved an accuracy of 91.6% for symbols recognition. Jayaraman et al. (2007) proposed a modular approach to recognize the online handwritten strokes for Telugu script. The authors identified 253 unique strokes, which are used to write the characters of Telugu script. Based on the relative position of the handwritten stroke in a character in their work, the stroke set has been classified into three subsets, viz., baseline strokes, bottom strokes, and top strokes. Three separate SVM-based classifiers have been built to classify the strokes in respective subsets. They have compared the SVM-based classifier results with the results computed by HMM-based classifier. Their experimental results show that the SVM-based online handwritten stroke classification performs better in term of accuracy as compared to HMM-based stroke classification. Their recognition system achieved an average stroke recognition accuracy of 87.7% for baseline strokes, 93.1% for bottom strokes, and 92.8% for top strokes. Prasanth et al. (2007) discussed the online handwritten Tamil and Telugu script recognition using elastic matching with local features. They have used Dynamic Time Wrapping (DTW) approach with four different feature sets, namely, x -, y -coordinate features; Shape Context (SC) and Tangent Angle (TA) features; Generalized Shape Context (GSC) features; and combination of normalized x -, y -coordinates, first and second derivatives, and curvature features. They have employed the Nearest Neighborhood classifier with DTW distance in their work. To test the performance of their recognition system, they considered 29,174 symbol samples for training and 9215 samples for testing. They have achieved an accuracy of 87.2% in their work. They have also reported the speed of recognition in this work. as 0.18 s per symbol. Figure 17 represents a subset of online handwritten Telugu characters samples.

Fig. 17 Telugu handwritten character samples, a subset taken from collected dataset



Rajkumar et al. (2012) presented Ternary Search Tree (TST) and Support Vector Machine (SVM) based online handwritten character recognition for Telugu script. In their work, the handwritten strokes used to write a Telugu character are considered in three vertical regions. The stroke recognition task is performed by using the SVM classifier and the separate SVM classifiers were trained to recognize the strokes in each region. They have collected character-level data from 100 writers involving 235 stroke classes. The collected data is pre-processed by applying the pre-processing transformations, namely, size normalization, smoothening, interpolation, and resampling. The features extracted include pre-processed x - and y -coordinate points, FFT, Hilbert transform (logarithm of the spectral density), and Wavelet transform. Their proposed approach achieved an overall stroke recognition accuracy of 89.6% for classification of 18,588 strokes by using Ternary Search Tree scheme and achieved an accuracy of 96.7% for classification of 14,037 strokes by using SVM-based classifiers. Thereafter, Chakravarthy and Team (2013) presented an online handwritten recognition system for Telugu characters using Support Vector Machine and Convolution Neural Networks. In their character recognition process, the strokes are recognized using Support Vector Machine and the character is recognized based on trained main rules using SVM. They have experimented a total of 20,654 handwritten strokes to test their system. On the other hand, they have also experimented the offline recognition of character approach by using CNNs and autoencoders. In these approaches, they have used four hidden layer CNN approach for vowel and consonants recognition. Their proposed online handwriting recognition system achieved an accuracy of 95.4% by using the ensemble classifier approach, which combines four networks (each of two hidden layers) selected among 10 different networks.

Table 3 summarizes the work done for Indic and non-Indic scripts, wherein, various features and methodologies used for online handwriting recognition system, along with their results have been presented.

4 Comparative study of classification techniques

Once a feature extraction or classification procedure finds the proper representation, a classifier can be designed using different number of approaches. However, it is very difficult to identify the best classifier that classify the input featured data to a specific category accurately. After reviewing the literature on online handwriting recognition system, we found that most of the existing literature focuses on extracting the structural, statistical, and linguistic features at stroke-, numeral-, character-, akshara-, and word-level as illustrated in Table 4. Most of the research work on online handwritten script recognition have been reported using the traditional classifiers and handcrafted features. However, in the recent literature, deep neural network based approaches have achieved the great superiority over the traditional statistical and structural feature based classifiers (Bhunia et al. 2020; Mukherjee et al. 2019; Budhouliya et al. 2020). Recently, the authors in Bhunia et al. (2020) proposed the offline-online multi-modal deep network framework to identify the 11 Indic-scripts. They have extracted more robust sequential features from the handwritten data using Convolution-LSTM Network and captured contextual information for better performance. They conclude that, in future the recognition of offline and online handwriting would be more promising using a single deep neural network model. Moreover, in Mukherjee et al. (2019) the authors proposed a novel approach to recognized the handwritten words wherein, they have extracted two different features namely, spatial (captured

Table 3 Summary of existing OHWR systems for Indic and non-Indic scripts

Script	References	Features	Methodology/tool	Results
Assamese	Reddy et al. (2012b)	x -, y -coordinates, vertical projection profile, and horizontal projection profile, DCT, chain code histogram and pixel level features, and vector quantization	HMM and vector quantization (VQ) classifiers	99.3% Numerals recognition accuracy
	Reddy et al. (2012a)	Pre-processed x -, y -coordinates, 1st and 2nd derivative of x - and y -coordinate points	HMM classifier	96.0% average handwritten numerals recognition accuracy
	Prasanna et al. (2013)	Combination of pre-processed x -, y -coordinates and 1st and 2nd derivative of x - and y -coordinate points	HMM classifier	95.2% Numerals recognition accuracy
	Baruah and Hazarika (2015)	Pre-processed x -, y -coordinates, directional, geometrical, and statistical features	SVM classifier	99.1% for Assamese numerals recognition accuracy and 81.2% for characters
	Choudhury et al. (2015)	Pre-processed x -, y -coordinates and 1st and 2nd derivative of x - and y -coordinate points	HMM and SVM classifiers	96.2% accuracy achieved for Assamese characters
Bangla	Garain et al. (2002)	Angle variation, Euclidean distance, and 8-directional coding	Template matching	96.3% Accuracy on 2440 basic Bangla characters
	Bhattacharya et al. (2007)	Direction codes	MLP classifier	83.6% accuracy on Bangla characters test dataset
	Parui et al. (2008)	6-Scalar features	HMM classifier	84.6% stroke classification accuracy
	Bhattacharya et al. (2008)	Histogram of direction codes	MQDF classifier	82.3% word level recognition accuracy
	Bandyopadhyay and Chakraborty (2009)	8-Directional chain coding, 12 shape features such as bump points and critical points	DTW classifier	96.6% character recognition accuracy
	Fink et al. (2010)	8-Scalar features for each sub-stroke	HMM classifier	93.1% context-dependent sub-word unit accuracy

Table 3 (continued)

Script	References	Features	Methodology/tool	Results
	Mohiuddin et al. (2011)	Sub-stroke level feature extraction	MLP and SVM classifiers	88.8% word-level (on 50 city names) recognition accuracy
	Biswas et al. (2012)	Center of gravity, Dirichlet distribution and trivariate normal distribution	HMM classifier	91.9% Character recognition accuracy
	Chowdhury et al. (2013)	Directional and positional information	Distance function based on Levenshtein distance metric	98.4% Accuracy for Bangla numerals
	Samanta et al. (2014)	Circular feature, linear feature, and combination of the two	HMM classifier	89.7% Word recognition accuracy
	Bhattacharya et al. (2018a)	Mod-NPen features (includes direction, curvature, aspect ratio, curliness, slope and context maps features)	HMM	93.5% accuracy achieved for unconstrained handwritten Bangla words
	Bhattacharya et al. (2018b)	Sub-stroke-wise relative features	HMM	94.1% accuracy achieved for cursively handwritten Bangla words
	Mukherjee et al. (2019)	Spatial, temporal, and PHOG features	Deep LSTM	81.5% accuracy achieved for Bangla handwritten words
Devanagari	Connell et al. (2000)	$dx, y, \sin(\theta)$, curvature and orientation, stroke direction histogram, dx and dy from beginning to end of each stroke, and center of gravity	HMM and NN classifiers	86.5% character recognition accuracy with no rejects
	Garain et al. (2002)	Angle variation information, Euclidean distance between two adjacent points, 8-directional coding information	Template matching approach	97.9% character recognition accuracy
	Namboodiri and Jain (2004)	11 Different spatial and temporal features	K-NN, neural net, and SVM-based classifiers	87.1% word-level accuracy with five-fold cross-validation on the data set of 13,379 words

Table 3 (continued)

Script	References	Features	Methodology/tool	Results
	Joshi et al. (2005)	Syntactic and structural features	Feature based recognition algorithm	On the test data set of 1487 Devanagari characters, 88.9% accuracy achieved when shirorekha was removed and 87.1% accuracy achieved without removing the shirorekha
	Swethalakshmi et al. (2006)	Pre-processed x -, y -coordinates, stroke length	SVMs and HMMs classifiers	96.7% Character recognition accuracy for 46 stroke-classes and 97.3% recognition accuracy for 82 stroke-classes
	Mondal et al. (2010)	Chain code histogram and point-float features	Nearest neighbour (NN), multi-layer perceptron (MLP), and Hidden Markov Model (HMM) classifiers	95.2% character recognition accuracy
	Bharath and Madhvanath (2012)	Aspect ratio, height fraction, and width fraction	HMM classifier, Lexicon-driven and lexicon-free strategies, Viterbi decoding strategy	93.4% and 87.1% accuracies have been achieved for 1000 and 20,000 size lexicons, respectively with combined lexicon-driven and lexicon-free strategy
	Belhe et al. (2012)	Histogram of oriented gradients (HOG)	HMM classifier and symbol tree approach	89.0% accuracy for isolated Devanagari words
	Mehrotra et al. (2013)	Pre-processed x -, y -coordinates	Convolutional neural network (CNN) classifier	98.2% character recognition accuracy
	Bhattacharya et al. (2018b)	Sub-stroke-wise relative features	HMM	88.1% accuracy achieved for cursively handwritten Devanagari words
	Mukherjee et al. (2019)	Spatial, temporal, PHOG features	Deep LSTM	82.5% accuracy achieved for Devanagari handwritten words
Gurmukhi	Sharma et al. (2008) Sharma et al. (2009a)	Pre-processed x -, y -coordinates High-level features such as loop, headline, straight line, and dot	Elastic matching Features based stroke recognition algorithm	90.1% character recognition accuracy Average stroke recognition accuracy of 94.6% with pre-processing and an accuracy of 88.8% achieved without pre-processing

Table 3 (continued)

Script	References	Features	Methodology/tool	Results
	Sharma et al. (2009c)	High-level features such as loop, headline, straight line, and dot	Elastic matching technique	81.1% word recognition accuracy
	Kumar and Sharma (2013)	Pre-processed x -, y -coordinates	SVM classifier and rule-Based approach	95.6% character recognition accuracy
	Sharma et al. (2013)	Pre-processed x -, y -coordinates	SVM classifier and rule-based approach	96.8% character recognition accuracy
	Kumar et al. (2015)	Pre-processed x -, y -coordinates	SVM classifier	80.4% Gurmukhi Akshara recognition accuracy
	Singh and Sachan (2015)	Statistical features including headline, straight line, and dots features	Nearest neighbor and SVM classifiers	86.9% word recognition accuracy
	Verma and Sharma (2015)	Zone based features including normalized, diagonal, directional, parabola based curve fitting and power curve based	SVM classifier	92.1% character recognition accuracy
	Verma and Sharma (2016)	Normalized x -, y -coordinates, region-based, curvature, curvature feature-based classes, and direction features	SVM classifier	96.7% character recognition accuracy for 35 basic characters
	Singh et al. (2016)	Chain codes features	k -means clustering, HMM classifier	Maximum 87.1% accuracy achieved for recognizing the middle-zone strokes
	Verma and Sharma (2017b)	Pre-processed x -, y -coordinates	SVM classifier	74.8% character recognition accuracy
	Verma and Sharma (2017a)	Pre-processed x -, y -coordinates	HMM classifier	An accuracy of 88.4% achieved for Gurmukhi consonants and their combinations with vowels. An accuracy of 97.1% achieved for basic Gurmukhi consonants
	Dahake et al. (2017)	Pre-processed x -, y -coordinates	SVM classifier	An accuracy of 91.0% achieved for Gurmukhi word segmentation

Table 3 (continued)

Script	References	Features	Methodology/tool	Results
Kannada	Singh et al. (2018a)	Pre-processed x - y coordinates, discrete Fourier transform, and directional features	SVM classifier	97.1% character recognition accuracy and 87.23% for character combinations with <i>Vowels, Nasals</i> and both
	Singh et al. (2018b)	Pre-processed x - y coordinates	SVM classifier, FSA	97.3% character recognition accuracy
	Budhouliya et al. (2020)	Online to offline converted images	CNN classifier	96.8% character recognition accuracy
	Prasad et al. (2009)	Normalized horizontal and vertical coordinates (a_i and b_j), normalized trajectory features (r_i and θ_j), and normalized deviation features (a_{di} and b_{dj})	k -NN classifier	81.0% character recognition accuracy
	Kunwar et al. (2010)	High level structural features, two estimates from 1st derivative	Statistical dynamic space warping (SDSW) classifier	Achieved 80.0% Unicode character level accuracy
	Murthy and Ramakrishnan (2011)	Pre-processed x -, y -coordinates, quantized slope, dominant points, and quartile features	Prototype-based DTW classifier, PCA	92.65% stroke classification accuracy
Malayalam	Rampalli and Ramakrishnan (2011)	Pre-processed x -, y -coordinates, pen direction angle, and 1st and 2st derivatives of x - and y -coordinates. Directional distance distribution (DDD), distance of Nearest Stroke pixels (NSP), Transition count (TC), and Projection profiles (PP)	SVM classifier	92.3% character recognition accuracy
	Ramakrishnan and Shashidhar (2013)	Pre-processed x -, y -coordinates	Attention Feed-based Segmentation and SVM classifier	Achieved an average accuracy of 94.3%, 62.0%, and 18.2% for word segmentation, symbol recognition, and word recognition, respectively
	Gowri Shankar and Chakravarthy (2003)	Shape features	String matching technique	90.8% stroke recognition accuracy

Table 3 (continued)

Script	References	Features	Methodology/tool	Results
	Arora and Nambodiri (2010)	Raw x - and y -coordinates, strokes movement, direction and curvature of stroke, length of stroke, aspect ratio, and area of stroke	HMM classifier and Viterbi-decoding process	78.1% word recognition accuracy
	Primekumar and Idiculla (2011)	Features from normalized x -, y -coordinates, wavelet transformation of resulted normalized feature vector	Wavelet transform and Simplified Fuzzy ARTMAP	97.8% character recognition accuracy using "haar" wavelet
	Indhu and Bhadran (2012)	Structural and directional features	Simplified Fuzzy ARTMAP with artificial neural network	98.3% average stroke recognition accuracy
	Nambodiri and Team (2013)	Inverse curvature re-sampling, circle based representation of strokes, and histogram of oriented gradients	SVM-DDAG classifier and LeNet-5 convolutional neural network	Achieved an average of 96.3% stroke recognition accuracy
Tamil	Aparna et al. (2004)	18 shape features such as dot, line terminal, bumps, cusp etc.	Soft-matching approach	82.8% stroke recognition accuracy
	Deepu et al. (2004)	Pre-processed x -, y -coordinates	PCA and NN	90.8% character recognition accuracy
	Joshi et al. (2004)	Pre-processed x -, y -coordinates, quantized slope, and coordinates of Dominant points	Elasting matching technique	95.9% character accuracy achieved by using the combination of rigid and elastic matching approach, and 94.8% accuracy achieved using dominant point based two-stage scheme
	Bharath and Madhvanath (2007)	Angle feature and curvature of trajectory points feature	HMM classifier	Achieved an average accuracy of 98.0% for 1000 size lexicons and 92.2% for 20,000 size lexicons
	Sundaram and Ramakrishnan (2008)	Radial distance and polar angle, radial distance from quartile mean and polynomial fit	Two dimensional PCA	81.1% character recognition accuracy
	Sundaram and Ramakrishnan (2009)	Normalized x -, y -coordinates, distance and angle features, radial distance and polar angle, quadratic fit, and autoregressive coefficient	Two dimensional PCA, 1	86.5% character recognition accuracy

Table 3 (continued)

Script	References	Features	Methodology/tool	Results
Telugu	Sundaram and Ramakrishnan (2013)	Pre-processing x -, y -coordinate features	AFS and DOCS strategies, SVM classifier	Word recognition accuracy of 50.9% and 64.9% achieved using DOCS and AFS strategies, respectively
	Ramakrishnan and Urala (2013)	Local features (pre-processed x -, y -coordinates), global features (DCT, DFT), and combination of both local and global features	SVM classifier	Achieved 95.9% character recognition accuracy using the combined features approach
	Sundaram and Ramakrishnan (2015)	Pre-processing x -, y -coordinates	SVM classifier, DTW approach	The accuracies of 93.0% and 81.6% achieved for symbol-level and word-level, respectively
	Swethalakshmi et al. (2006)	Pre-processed x -, y -coordinates, stroke length	SVM classifiers	83.1% average stroke recognition accuracy achieved by the classifier trained using 33,726 samples
	Babu et al. (2007)	11 time-domain features and 8 frequency-domain features	HMM classifier	91.6% symbols recognition accuracy
	Jayaraman et al. (2007)	Shape features	HMMs and SVMs classifiers	Achieved 87.7% accuracy for baseline strokes, 93.1% for bottom strokes, and 92.8% for top strokes
	Prasanth et al. (2007)	x -, y -coordinate features, Shape context (SC) and Tangent angle (TA) features, Generalized Shape Context (GSC), and combination of x -, y -coordinates, 1st and 2nd derivatives and curvature features	DTW classifier	29,174 samples used in training. 87.2% symbol recognition accuracy on 9215 test samples
	Rajkumar et al. (2012)	Pre-processed x -, y -coordinates, Fourier transformation, FFT, Hilbert transform, and wavelet features	Ternary Search Tree (TST) and SVM classifier	89.6% overall stroke recognition accuracy for classifying 18,588 strokes using Ternary Search Tree scheme and 96.7% for classifying 14,037 strokes using SVM-based classifier
	Chakravarthy and Team (2013)	FFT, DFT, and shape based feature	SVM and CNN classifiers	95.4% character recognition accuracy

Table 3 (continued)

Script	References	Features	Methodology/tool	Results
Roman	Tappert et al. (1990)	Shape based feature: smoothing, wild point correction, de-hooking, dot reduction, size-normalization, and stroke length normalization	Elastic matching, Decision Tree	95.0% character recognition accuracy
	Veltman and Prasad (1994)	Spatial features, dot detection, de-hooking, rotation independency, size-normalization, and quantization	HMM	Achieved an average error rate of 6.9% for character recognition
	Tan et al. (2009)	Horizontal and vertical inter-stroke direction, horizontal and vertical stroke direction, average stroke length, stroke density and the reverse direction	Information Retrieval (IR) model	84.0% accuracy achieved for Roman script recognition
Arabic	Almuallim and Yamaguchi (1987)	Cross point, branch point, line start, line end, stroke length, frame of the stroke, and connection point with previous stroke	String Matching technique	81.3% word recognition accuracy
	El-Wakil and Shoukry (1989)	Number of dots (NDOT), relative position of dots (PDDOT), Number of secondary strokes (STROKE), and Slope of secondary stroke (SS)	Template matching technique with tree structure and k -NN classifier	84.0% accuracy achieved for characters
Chinese	Beigi et al. (1994)	Stroke based extreme velocities and geometric features	HMM classifier	An accuracy of 93.1% achieved for Arabic digits
	Liu et al. (2004)	Noise elimination and shape normalization	Structural matching, HMM, and statistical classification	98.0% accuracy achieved for Chinese regular characters and 90.0% for fluent-regular character
	Bai and Huo (2005)	8-Directional features	Single prototype classifier and 1-NN classifier	The highest stroke classification accuracy of 99.8% achieved

Table 3 (continued)

Script	References	Features	Methodology/tool	Results
	Liu et al. (2013)	Gradient directional features, directional features	MQDF, NPC,DFE and DLQDF.	Accuracies achieved is 92.1% and 94.9% on the HWDDB1.1 (offline) and OLHWDB1.1 (online) datasets, respectively
Japanese	Takahashi et al. (1997)	Directional features	HMM	95.4% accuracy achieved for Kanji (Japanese) characters recognition
	Jäger et al. (2003)	Directional and histogram features	Dynamic programming-matching, HMM, and Tree based character recognition methods	90.0% accuracy achieved for Kuchibue character database
	Liu and Zhou (2006)	8-chaincode direction and directional decomposition of a line segment features	Euclidean distance and modified quadratic discriminant function (MQFD2)	91.0% accuracy achieved using MQFD2
	Zhou et al. (2007)	Geometric features	Path evaluation approach based on geometric and linguistic context	94.7% accuracy achieved using unary geometry
Persian	Baghshah et al. (2006)	Start2End_Direction, Start2COG_Direction, End2COG_Direction, Straightness, Horizontal_Motion, Vertical_Motion, Curvature_Side, and Aspect ratio	Fuzzy Learning Vector Quantization (FLVQ)	95.0% accuracy achieved to recognize the Persian characters
	Halavati and Shouraki (2007)	Segment type, segment direction, curvature direction, and segment length	Elastic pattern matching and fuzzy modeling	Achieved an accuracy rate of 78.0% without using dictionary and 96.0% with using dictionary
	Izadi et al. (2008)	Wavelet-based smoothing, Convex Curve Sectors (CCSs)	Dynamic time warping (DTW)	An accuracy rate of 89.4% achieved for 2-letter word and 85.0% for 3-letter word
	Ghods et al. (2013)	Delayed strokes including dots and small signs	HMM	An accuracy rate of 85.2% achieved for online handwritten Persian words
Thai	Budsayaplakorn et al. (2003)	Angle and sub-stroke features	HMM and Fuzzy logic	92.1% accuracy achieved for Thai characters using HMM+Fuzzy and 89.3% using only HMM

Table 3 (continued)

Script	References	Features	Methodology/tool	Results
	Sanguansat et al. (2004)	Angle and sub-stroke features	Structural matching, HMM, SVM, Likelihood-Ratio (LR) and Symmetrical Likelihood-Ratio score-space (SLR)	Accuracy achieved 96.2% for writer-independent and 92.5% for writer-independent using HMM+SVM+SLR for Thai character recognition
	Karnchanapusakij et al. (2009)	Head, end, loop, twist, piecewise curve, and straight line features	Linear interpolation approach	90.9% accuracy achieved for Thai characters recognition

Table 4 Classification methodology-wise comparative study of recognition results at Stroke-, Character-, Akshara-, and Word-level for Indic and non-Indic scripts

Methodology	Script (language)	References	Classifier(s)	Recognition level	Accuracy (%)
<i>Indic scripts</i>					
Statistical	Assamese (Assamese)	Reddy et al. (2012a)	HMM	Numeral	99.3
		Reddy et al. (2012b)	HMM	Numeral	96.0
		Prasanna et al. (2013)	HMM	Numeral	95.2
		Baruah and Hazarika (2015)	SVM	Character	99.1
		Choudhury et al. (2015)	HMM, SVM	Character	96.2
	Bangla (Bengali)	Parui et al. (2008)	HMM	Character	84.6
		Bhattacharya et al. (2008)	MQFD	Word	82.3
		Fink et al. (2010)	HMM	Word	93.1
		Biswas et al. (2012)	HMM	Character	91.9
		Samanta et al. (2014)	HMM	Word	89.7
	Devanagari (Hindi)	Bhattacharya et al. (2018b)	HMM	Word	94.1
		Bhattacharya et al. (2018a)	HMM	Word	94.5
		Swehalakshmi et al. (2006)	SVM, HMM	Character	97.3
		Bharath and Madhvanath (2012)	HMM	Word	93.4
		Belhe et al. (2012)	HMM	Word	89.0
Gurmukhi (Punjabi)	Bhattacharya et al. (2018b)	HMM	Word	88.1	
	Kumar et al. (2015)	SVM	Akshara	80.4	
	Verma and Sharma (2016)	SVM	Character	96.7	
	Singh et al. (2016)	HMM, <i>k</i> -NN	Stroke	87.1	
	Verma and Sharma (2017b)	SVM	Character	74.8	
Kannada (Kannada)	Verma and Sharma (2017a)	HMM	Character	97.1	
	Dahake et al. (2017)	SVM	Word	91.0	
	Prasad et al. (2009)	<i>k</i> -NN	Character	81.0	
	Kunwar et al. (2010)	SDSW	Character	80.0	
	Rampalli and Ramakrishnan (2011)	SVM	Character	92.3.0	

Table 4 (continued)

Methodology	Script (language)	References	Classifier(s)	Recognition level	Accuracy (%)
Structural	Malayalam (Malayalam)	Arora and Nambodiri (2010)	HMM	Word	78.1
	Tamil (Tamil)	Bharath and Madhavanath (2007)	HMM	Word	98.0
		Sundaram and Ramakrishnan (2008)	2D-PCA	Character	81.0
		Ramakrishnan and Urala (2013)	SVM	Character	95.9
	Telugu (Telugu)	Swethalakshmi et al. (2006)	SVM	Stroke	83.1
		Babu et al. (2007)	HMM	Symbol	91.6
		Jayaraman et al. (2007)	HMM, SVM	Stroke	3.1
		Rajkumar et al. (2012)	SVM, TST	Stroke	89.6
		Garain et al. (2002)	Template matching	Character	96.3
		Bandyopadhyay and Chakraborty (2009)	DTW	Character	96.6
Neural network	Bangla (Bengali)	Chowdhury et al. (2013)	Template matching	Numeral	98.4
	Devanagari (Hindi)	Garain et al. (2002)	Template matching	Character	97.9
		Joshi et al. (2005)	Template matching	Character	88.9
	Gurmukhi (Punjabi)	Sharma et al. (2008)	Elastic matching	Character	90.1
		Sharma et al. (2009b)	Elastic matching	Word	81.1
	Malayalam (Malayalam)	Gowri Shankar and Chakravarthy (2003)	String matching	Stroke	90.8
	Tamil (Tamil)	Aparna et al. (2004)	Soft matching	Stroke	82.8
		Joshi et al. (2004)	Elastic matching	Character	95.9
		Prasanth et al. (2007)	DTW	Symbol	87.2
		Bhattacharya et al. (2007)	MLP	Character	83.6
Mixed	Telugu (Telugu)	Mukherjee et al. (2019)	Deep LSTM	Word	81.5
	Bangla (Bengali)	Mehrotra et al. (2013)	CNN	Character	98.2
	Devanagari (Hindi)	Mukherjee et al. (2019)	Deep LSTM	Word	82.5
	Gurmukhi (Punjabi)	Budhouliya et al. (2020)	CNN	Character	96.8
	Bangla (Bengali)	Mohiuddin et al. (2011)	MLP, SVM	Word	88.8
Devanagari (Hindi)	Connell et al. (2000)	HMM, NN	Character	86.5	

Table 4 (continued)

Methodology	Script (language)	References	Classifier(s)	Recognition level	Accuracy (%)		
Non-Indic scripts	Statistical	Namboodiri and Jain (2004) Mondal et al. (2010) Kumar and Sharma (2013) Singh and Sachan (2015) Singh et al. (2018a) Singh et al. (2018b) Murthy and Ramkrishnan (2011) Primekumar and Idiculla (2011) Indhu and Bhadrani (2012) Namboodiri and Team (2013) Deepu et al. (2004) Sundaram and Ramkrishnan (2009) Sundaram and Ramkrishnan (2013) Sundaram and Ramkrishnan (2015) Chakravarthy and Team (2013)	k-NN, SVM, NN NN, MLP, HMM SVM, Rule-based NN, SVM SVM, Rule-based SVM, FSA PCA with NN, DTW Wavelet transform, Fuzzy ARTMAP Fuzzy ARTMAP, ANN SVM-DDAG, LeNet-5, CNN PCA, NN 2D-PCA, NN 2D-PCA, DOCS, SVM SVM, DTW SVM, CNN	Word Character Character Word Character Character, Akshara Akshara Character Stroke Stroke Character Character Word Symbol, Word Character	87.1 95.2 95.6 86.9 97.1 97.3, 87.2 92.65 97.8 98.3 96.3 90.8 86.5 64.9 93.0, 81.6 95.4		
			Structural	Veltman and Prasad (1994) Tan et al. (2009) Beigi et al. (1994) Liu et al. (2013) Takahashi et al. (1997) Liu and Zhou (2006) Ghods et al. (2013) Tappert et al. (1990) Almuallim and Yamaguchi (1987)	HMM IR model HMM MQDF, NPC, DFE, and DLQDF HMM MQFD2 HMM Elastic matching String matching	Character Script Digits Character Character Character Word Character Word	93.1 84.0 93.1 94.9 95.4 91.0 85.2 95.0 81.3

Table 4 (continued)

Methodology	Script (language)	References	Classifier(s)	Recognition level	Accuracy (%)
Neural Network	Japanese (Japanese)	Zhou et al. (2007)	Path Evaluation approach	Character	94.7
	Thai (Thai)	Karnchanapuskijj et al. (2009)	Linear Interpolation approach	Character	90.9
	Persian (Persian)	Izadi et al. (2008)	Dynamic Time Warping (DTW)	Word	89.4
	Chinese (Chinese)	Bai and Huo (2005)	Neural Network	Stroke	99.8
Mixed		Zhang et al. (2017)	CNN	Character	99.8
	Arabic (Arabic)	El-Wakil and Shoukry (1989)	Template matching, k -NN	Character	84.0
	Chinese (Chinese)	Liu et al. (2004)	Structural Matching, HMM	Character	98.0
	Japanese (Japanese)	Jäger et al. (2003)	Dynamic Programming matching, HMM, Tree	Character	90.0
	Thai (Thai)	Budsayaplakorn et al. (2003)	HMM, Fuzzy logic	Character	92.1
		Sanguansat et al. (2004)	Structural matching, HMM, SVM, SLR	Character	96.2
	Persian (Persian)	Baghshah et al. (2006)	FLQV	Character	95.0
		Halavati and Shouraki (2007)	Elastic Pattern matching, Fuzzy Modeling	Word	96.0

from offline handwritten data) and temporal (pseudo-online) from offline handwritten word image. Furthermore, Pyramid Histogram Oriented Gradient (PHOG) features are extracted without any zone segmentation. They have also reported the recognition of handwritten words using deep LSTM Network. To evaluate the performance of their word recognition system, they considered two Indic-scripts namely, Devanagari and Bangla. According to them, the combination of offline and online classifiers, significantly outperform in order to recognize the handwritten words. Moreover, the same idea (offline and online classifiers based classification) can also be used for other scripting languages to improve their recognition rate. Also, in a recent article Budhouliya et al. (2020), the authors presented the online handwritten Gurmukhi character recognition using the Convolution Neural Network. In their work they increase the online handwritten stroke images data using the Generative Adversarial Network (GAN), a data augmentation technique. Their experiments achieved an optimal Gurmukhi character recognition accuracy of 96.8%.

In the last two decades, it has been observed that a number of online handwriting recognition systems have been developed. Among those systems, most of them have used statistical/stochastic methods (i.e., Hidden Markov Models), network based methods (i.e., Neural Networks), and structural/syntactical methods (i.e., Elastic Matching). It has been noticed that the major drawback of all the aforementioned methods is their high-dependability on small writing perturbations that tends to incorrect recognition results, when working on huge variety of handwriting styles. To overcome the limitations or drawbacks of these methods, the online handwriting recognition system should possess: fast response and flexible. To develop such system, (1) the knowledge base must be small and robust, and (2) new variation in handwriting styles should be handled by flexible prototypes, that contain the handwritten character/word information widely (Malaviya and Klette 1996).

Moreover, in recent years' studies, most of the online handwriting recognition work is done using the Deep Neural Network based approaches. The online handwriting samples are converted into offline images and then Deep Neural Network approaches (LSTM, GAN, DLSTM) are applied to obtain the promising recognition accuracy (Mukherjee et al. 2019). However, the recognition performance of deep neural network based systems is appreciable but there are still some limitations of these methodologies i.e., required large amount of dataset for training the classifier, high performance CPU or GPU, takes long time to train the classifier etc. These limitations can be overcome by using other advanced approaches, i.e., Generative Adversarial Network (GAN), used to extend or augment the small/primary dataset at large scale.

5 Datasets availability

In the recent past, the research work in the direction of online handwriting recognition systems has received much attention. To the best of our knowledge, there exists no such a common platform to make available the standard datasets (online handwritten) for Indic scripts. Here, in Table 5 we have collected various publicly available datasets of different scripting language with their names. Moreover, the information about the level(s) (i.e., stroke-, numeral-, character-, and word-level), name, and scripting language of respective dataset is available in this table. Here, the scripting languages (Indic and non-Indic), those handwritten dataset is publicly available and included in this study are Roman, Arabic, Chinese, Japanese, Assamese, Bangla, Devanagari, Gurmukhi, Kannada, Tamil, and Telugu. This is worth mentioned here that there are some scripting languages, those writing

Table 5 List of publicly available Datasets for Indic and non-Indic scripts

References	Scripting language	Level of data (Stroke/numeral/character/word)	Dataset name	Resources/URL(s)
Guyon et al. (1994)	Roman	Characters, words	UNIPEN collection train_r01_v07	http://unipen.org/ , https://www.ai.rug.nl/~lambert/unipen/1cdar-03-compe-tition/
Mahmoud et al. (2018)	Arabic	Words	Online-KHATT	http://onlinekhatt.ideas2serve.net/
Liu et al. (2013)	Chinese	Characters, words	CASIA-OLHWDB	http://www.nipr.ia.ac.cn/databases/handwriting/Download.html
Matsumoto et al. (2001)	Japanese/Kanji	Characters	Japanese character database	http://web.tuat.ac.jp/~makagawa/databse/index.html , http://lipitk.sourceforge.net/resources.htm
Baruah and Hazarika (2015)	Assamese	Characters	Artificial characters	http://mlr.cs.umass.edu/ml/datasets/Online+Handwritten+Assamese+Characters+Dataset
Mondal et al. (2010)	Bangla	Numerals, characters	ISI Bangla character database	http://lipitk.sourceforge.net/resources.htm
Mondal et al. (2010)	Devanagari	Characters, word	Isolated handwritten Devanagari character dataset, isolated handwritten Devanagari word dataset	http://lipitk.sourceforge.net/hpl-datasets.htm
Singh et al. (2019)	Gurmukhi	Strokes, numerals, characters, Aksharas, words	OHWR-GScript	https://sites.google.com/view/ohwr-gurmukhi-script/
Ramakrishnan and Shashidhar (2013)	Kannada	Strokes, characters, words	MILE Lab Kannada symbols dataset, MILE Lab Kannada word dataset	http://mile.ee.iisc.ac.in
Mondal et al. (2010)	Tamil	Characters, words	Isolated handwritten Tamil character dataset, isolated handwritten Tamil word dataset	http://lipitk.sourceforge.net/hpl-datasets.htm , http://mile.ee.iisc.ac.in/Downloads/ohr.zip
Mondal et al. (2010)	Telugu	Characters	Isolated handwritten Telugu character dataset	http://lipitk.sourceforge.net/hpl-datasets.htm , http://mile.ee.iisc.ac.in/downloads/hpl.zip

structure is similar to each other like, Devanagari and Gurmukhi, Devanagari and Bangla. Therefore, the feature extraction techniques and classification methodologies used in these scripting languages may help the researchers to improve their recognition accuracy. We believe that this description of common availability of the online handwritten datasets of many scripting languages have the potential to support future research in this area.

6 Future scope of the study

The research work reported on online handwriting recognition systems for non-Indic and Indic scripts may be extended in several directions. Some of them are discussed below.

- (1) *OHR for official, cursive handwriting*: In Indic script recognition, most of the work have been reported on isolated character recognition and good-quality handwritten character recognition. A very little amount of work have been carried-out for the recognition of cursive handwritten words for various Indic-scripts. Experiments should be made to judge the effect of cursively written words as well as noise of various types, and take corrective measures.
- (2) *Integration of language model*: For online handwriting recognition task, we need to start from the atomic-level. Herein, strokes are considered as the atomic-levels. After combining the recognized strokes a character/word is formed. Although, some online handwriting recognizers are efficient in character recognition but, comparative less accurate in word recognition. Therefore, integration of language model in an online handwriting recognition system can improve the accuracy at word and sentence recognition level.
- (3) *Multi-script OHR development*: Since, India is a multi-lingual country, where various different scripting languages are used for communication. Therefore, It would be highly recommended to develop multi-script OHR system for intra-state communication. In multi-script OHR system development, identification of different script is the major challenge due to their complex structure. Some techniques to identify the script for offline character recognition systems have been reported (Pal and Chaudhuri 1997, 2003).
- (4) *Improvement of post processing algorithms*: An appreciable rate of accuracy have been reported by many researchers for recognizing the handwritten characters only. However, the existing post processing techniques are used to improve the recognition accuracy at word and sentence level (Kumar and Sharma 2013; Singh et al. 2018b). But still there is need to exploit more efficient post processing techniques to achieve the better recognition accuracy at word or sentence level.
- (5) *writer identification*: Online handwriting recognition system can be further extended for writer identification. Traditionally, handwriting analysis was used in the field of forensic document analysis by forensic expert in the detection of fraud, theft, and embezzlement cases. The potential applications of writer identification are forensic expert decision making systems, writer adaptation in mobile devices etc.
- (6) *Adaptation of new handwriting style*: The online handwriting recognition systems recognize the handwritten character or words by using the trained model. Wherein, fixed number of writer's handwriting samples are considered. These kind of systems are not able to include the new style of handwriting samples in the existing trained model. Such

kind of OHR system can be developed which can adapt the new handwriting styles. is being trained one time using the fixed number of samples

- (7) *Generation of Benchmarking datasets*: Many non-Indic script's online handwritten datasets are publically available for research work and competition (Messoud et al. 2012; Djeddi et al. 2016). An attempt can be made in this direction by creating a common platform and make the standard benchmark datasets of Indian scripting languages and make that publically available for research endeavors in the area of online handwriting recognition.

7 Conclusion

In this paper, a comprehensive review for the recognition of online handwritten numerals, characters, aksharas, and words for Indic and non-Indic scripts has been carried out. A brief discussion about the pre-processing steps, feature extraction techniques, classification tools and methodologies used, and various post-processing techniques applied for online handwriting recognition development have been carried out. Moreover, we have summarized the recognition work of the reviewed scripts in terms of feature extracted; accuracy achieved; and classification methodologies employed. Further, classification methodology wise comparative study of recognition results at Stroke-, Character-, Akshara-, and Word-level for Indic and non-Indic scripts have also been discussed. Eventually, an attempt to provide the resources for online handwritten datasets, almost all of the reviewed scripts have been made available in this present study. We believe that this review will strongly encourage the researchers in the area of online handwriting recognition.

References

- Almuallim H, Yamaguchi S (1987) A method of recognition of Arabic cursive handwriting. *IEEE Trans Pattern Anal Mach Intell* 5:715–722
- Alsallakh B, Safadi H (2006) Arapen: an Arabic online handwriting recognition system. In: *Information and communication technologies, 2006. ICTTA'06. 2nd, vol 1*. IEEE, pp 1844–1849
- Aparna K, Subramanian V, Kasirajan M, Prakash GV, Chakravarthy V, Madhvanath S (2004) Online handwriting recognition for Tamil. In: *Ninth international workshop on frontiers in handwriting recognition, 2004. IWFHR-9 2004*. IEEE, pp 438–443
- Arora A, Namboodiri A M (2010) A hybrid model for recognition of online handwriting in Indian scripts. In: *2010 International conference on frontiers in handwriting recognition (ICFHR)*. IEEE, pp 433–438
- Babu V J, Prasanth L, Sharma R R, Bharath A (2007) Hmm-based online handwriting recognition system for Telugu symbols. In: *Ninth international conference on document analysis and recognition, 2007. ICDAR 2007, vol 1*. IEEE, pp 63–67
- Baghshah MS, Shouraki SB, Kasaei S (2006) A novel fuzzy classifier using fuzzy lvq to recognize online Persian handwriting. In: *2006 2nd International conference on information and communication technologies, vol 1*. IEEE, pp 1878–1883
- Bahri H (1982) *Teach yourself Panjabi*. Panjabi University, Patiala
- Bai Z-L, Huo Q (2005) A study on the use of 8-directional features for online handwritten Chinese character recognition. In: *Eighth international conference on document analysis and recognition, 2005. Proceedings*. IEEE, pp 262–266
- Bandyopadhyay A, Chakraborty B (2009) Development of online handwriting recognition system: a case study with handwritten Bangla character. In: *World congress on nature and biologically inspired computing, 2009. NaBIC 2009*. IEEE, pp 514–519
- Baruah U, Hazarika SM (2015) A dataset of online handwritten assamese characters. *J Inf Process Syst* 11(3):325–341

- Beigi H S, Nathan K, Clary G J, Subrahmonia J (1994) Challenges of handwriting recognition in Farsi, Arabic and other languages with similar writing styles an on-line digit recognizer. In: Proceedings of the 2nd annual conference on technological advancements in developing countries, Columbia University, New York. Citeseer
- Belhe S, Chakravarthy S, Ramakrishnan A (2009) Xml standard for Indic online handwritten database. In: Proceedings of the international workshop on multilingual OCR. ACM, p 19
- Belhe S, Paulzagade C, Surve S, Jawanjal N, Mehrotra K, Motwani A (2010) Annotation tool and XML representation for online Indic data. In: 2010 International conference on frontiers in handwriting recognition (ICFHR). IEEE, pp 664–669
- Belhe S, Paulzagade C, Deshmukh A, Jetley S, Mehrotra K (2012) Hindi handwritten word recognition using hmm and symbol tree. In: Proceeding of the workshop on document analysis and recognition. ACM, pp 9–14
- Bharath A, Madhvanath S (2007) Hidden Markov models for online handwritten Tamil word recognition. In: ICDAR. IEEE, pp 506–510
- Bharath A, Madhvanath S (2012) Hmm-based lexicon-driven and lexicon-free word recognition for online handwritten Indic scripts. *IEEE Trans Pattern Anal Mach Intell* 34(4):670–682
- Bhattacharya U, Gupta BK, Parui S (2007) Direction code based features for recognition of online handwritten characters of Bangla. In: Ninth international conference on document analysis and recognition, 2007. ICDAR 2007, vol 1. IEEE, pp 58–62
- Bhattacharya U, Nigam A, Rawat Y, Parui S (2008) An analytic scheme for online handwritten Bangla cursive word recognition. In: Proceedings of the 11th ICFHR, pp 320–325
- Bhattacharya N, Pal U, Roy PP (2017) Cleaning of online Bangla free-form handwritten text. *ACM Trans Asian Low Resour Lang Inf Process (TALLIP)* 17(1):1–25
- Bhattacharya N, Roy PP, Pal U, Setua SK (2018a) Online Bangla handwritten word recognition. *Malays J Comput Sci* 31(4):300–310
- Bhattacharya N, Roy PP, Pal U (2018b) Sub-stroke-wise relative feature for online indic handwriting recognition. *ACM Trans Asian Low Resour Lang Inf Process (TALLIP)* 18(2):1–16
- Bhunia AK, Mukherjee S, Sain A, Bhunia AK, Roy PP, Pal U (2020) Indic handwritten script identification using offline-online multi-modal deep network. *Inf Fusion* 57:1–14
- Biadisy F, El-Sana J, Habash N (2006) Online Arabic handwriting recognition using hidden markov models. In: Tenth international workshop on frontiers in handwriting recognition. Suvisoft
- Biswas C, Bhattacharya U, Parui SK (2012) Hmm based online handwritten Bangla character recognition using Dirichlet distributions. In: 2012 International conference on frontiers in handwriting recognition (ICFHR). IEEE, pp 600–605
- Bousslama F, Amin A (1998) Pen-based recognition system of Arabic character utilizing structural and fuzzy techniques. In: 1998 Second international conference on knowledge-based intelligent electronic systems, 1998. Proceedings KES'98, vol 3. IEEE, pp 76–85
- Budhouliya R, SHARMA RK, Singh, H (2020) Recognition of online handwritten Gurmukhi strokes using convolutional neural networks. In: Proceedings of the 12th international conference on agents and artificial intelligence—volume 2: ICAART. INSTICC, SciTePress, pp 578–586
- Budsayaplakorn R, Asdornwised W, Jitapunkul S (2003) On-line Thai handwritten character recognition using hidden markov model and fuzzy logic. In: 2003 IEEE 13th workshop on neural networks for signal processing, 2003. NNSP'03. IEEE, pp 537–546
- Chakravarthy V Team (2013) Development of OHWR system for Telugu language. *VishwaBharat J Lang Technol* 39&40:23–54
- Choudhury H, Mandal S, Devnath S, Prasanna SM, Sundaram S (2015) Combining hmm and svm based stroke classifiers for online assamese handwritten character recognition. In: 2015 Annual IEEE India conference (INDICON). IEEE, pp 1–6
- Chowdhury SD, Bhattacharya U, Parui SK (2013) Online handwriting recognition using Levenshtein distance metric. In: 2013 12th International conference on document analysis and recognition (ICDAR). IEEE, pp 79–83
- Connell SD, Sinha R, Jain AK (2000) Recognition of unconstrained online Devanagari characters. In: 15th International conference on pattern recognition, 2000. Proceedings, vol 2. IEEE, pp 368–371
- Dahake D, Sharma R, Singh H (2017) On segmentation of words from online handwritten Gurmukhi sentences. In: 2017 2nd International conference on man and machine interfacing (MAMI). IEEE, pp 1–6
- Datta A (1984) A generalized formal approach for description and analysis of major Indian scripts. *IETE J Res* 30(6):155–161

- Deepu V, Madhvanath S, Ramakrishnan A (2004) Principal component analysis for online handwritten character recognition. In: Proceedings of the 17th international conference on pattern recognition, 2004. ICPR 2004, vol 2. IEEE, pp 327–330
- Djeddi C, Al-Maadeed S, Gattal A, Siddiqi I, Ennaji A, El Abed H (2016) ICFHR2016 competition on multi-script writer demographics classification using “QUWI” database. In: 2016 15th International conference on frontiers in handwriting recognition (ICFHR). IEEE, pp 602–606
- El-Abed H, Märgner V, Kherallah M, Alimi AM (2009) Icdar 2009 online Arabic handwriting recognition competition. In: 10th International conference on document analysis and recognition, 2009. ICDAR'09. IEEE, pp 1388–1392
- El-Wakil MS, Shoukry AA (1989) On-line recognition of handwritten isolated Arabic characters. *Pattern Recognit* 22(2):97–105
- Fink GA, Vajda S, Bhattacharya U, Parui SK, Chaudhuri BB (2010) Online Bangla word recognition using sub-stroke level features and hidden Markov models. In: 2010 International conference on frontiers in handwriting recognition (ICFHR). IEEE, pp 393–398
- Garain U, Chaudhuri B, Pal T (2002) Online handwritten Indian script recognition: a human motor function based framework. In: 16th International conference on pattern recognition, 2002. Proceedings, vol 3. IEEE, pp 164–167
- Ghosh V, Kabir E, Razzazi F (2013) Effect of delayed strokes on the recognition of online farsi handwriting. *Pattern Recognit Lett* 34(5):486–491
- Ghosh D, Dube T, Shivaprasad A (2010) Script recognition—a review. *IEEE Trans Pattern Anal Mach Intell* 32(12):2142–2161
- Ghosh R, Kumar P, Roy PP (2019) A Dempster–Shafer theory based classifier combination for online signature recognition and verification systems. *Int J Mach Learn Cybern* 10(9):2467–2482
- Gowri Shankar VA, Chakravarthy V (2003) Lekhak [mal]: a system for online recognition of handwritten Malayalam characters. In: National conference on communications, IIT, Madras
- Guyon I, Schomaker L, Plamondon R, Liberman M, Janet S (1994) Unipen project of on-line data exchange and recognizer benchmarks. In: Proceedings of the 12th IAPR international conference on pattern recognition, 1994. Vol 2-conference B: computer vision and image processing, vol 2. IEEE, pp 29–33
- Halavati R, Shouraki SB (2007) Recognition of Persian online handwriting using elastic fuzzy pattern recognition. *Int J Pattern Recognit Artif Intell* 21(03):491–513
- Indhu TR, Bhadrn VK (2012) Malayalam online handwriting recognition system: a simplified fuzzy artmap approach. In: 2012 Annual IEEE India conference (INDICON), pp 613–618
- Izadi S, Haji M, Suen CY (2008) A new segmentation algorithm for online handwritten word recognition in Persian script. In: Proceedings of the eleventh international conference on frontiers in handwriting recognition (CFHR 2008), pp 598–603
- Jaeger S, Manke S, Reichert J, Waibel A (2001) Online handwriting recognition: the npen++ recognizer. *Int J Doc Anal Recognit* 3(3):169–180
- Jäger S, Liu C-L, Nakagawa M (2003) The state of the art in Japanese online handwriting recognition compared to techniques in western handwriting recognition. *Doc Anal Recognit* 6(2):75–88
- Jain AK, Mao J, Mohiuddin KM (1996) Artificial neural networks: a tutorial. *Computer* 29(3):31–44
- Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intell* 22(1):4–37
- Jayaraman A, Sekhar CC, Chakravarthy VS (2007) Modular approach to recognition of strokes in Telugu script. In: Ninth international conference on document analysis and recognition (ICDAR 2007), vol 1. IEEE, pp 501–505
- Joshi N, Sita G, Ramakrishnan A, Madhvanath S (2004) Comparison of elastic matching algorithms for online Tamil handwritten character recognition. In: Ninth international workshop on frontiers in handwriting recognition, 2004. IWFHR-9 2004. IEEE, pp 444–449
- Joshi N, Sita G, Ramakrishnan A, Deepu V, Madhvanath S (2005) Machine recognition of online handwritten Devanagari characters. In: Eighth international conference on document analysis and recognition, 2005. Proceedings. IEEE, pp 1156–1160
- Karnchanapusakij C, Suwannakat P, Rakprasertsuk W, Dejdumrong N (2009) Online handwriting Thai character recognition. In: Sixth international conference on computer graphics, imaging and visualization, 2009. CGIV'09. IEEE, pp 323–328
- Kherallah M, Tagougui N, Alimi AM, El Abed H, Margner V (2011) Online Arabic handwriting recognition competition. In: 2011 International conference on document analysis and recognition (ICDAR). IEEE, pp 1454–1458
- Kumar A, Bhattacharya S (2010) Online Devanagari isolated character recognition for the iphone using hidden Markov models. In: Students' technology symposium (TechSym), 2010 IEEE. IEEE, pp 300–304

- Kumar R, Sharma RK (2013) An efficient post processing algorithm for online handwriting Gurmukhi character recognition using set theory. *Int J Pattern Recognit Artif Intell* 27(04):1353002
- Kumar R, Sharma RK, Sharma A (2015) Recognition of multi-stroke based online handwritten Gurmukhi Aksharas. *Proc Natl Acad Sci India Sect A Phys Sci* 85(1):159–168
- Kunwar R, Mohan P, Shashikiran K, Ramakrishnan A (2010) Unrestricted Kannada online handwritten Akshara recognition using sdtw. In: 2010 International conference on signal processing and communications (SPCOM). IEEE, pp 1–5
- Liu C-L, Zhou X-D (2006) Online Japanese character recognition using trajectory-based normalization and direction feature extraction. In: Tenth international workshop on frontiers in handwriting recognition. Suvisoft
- Liu C-L, Jaeger S, Nakagawa M (2004) Online recognition of Chinese characters: the state-of-the-art. *IEEE Trans Pattern Anal Mach Intell* 26(2):198–213
- Liu C-L, Yin F, Wang D-H, Wang Q-F (2013) Online and offline handwritten Chinese character recognition: benchmarking on new databases. *Pattern Recognit* 46(1):155–162
- Mahmoud SA, Luqman H, Al-Helali BM, BinMakhashen G, Parvez MT (2018) Online-khatt: an open-vocabulary database for Arabic online-text processing. *Open Cybern Syst J* 12(1):42–59
- Malaviya A, Klette R (1996) A fuzzy syntactic method for on-line handwriting recognition. In: International workshop on structural and syntactic pattern recognition. Springer, pp 381–392
- Mandal S, Prasanna SM, Sundaram S (2015) Curvature point based hmm state prediction for online handwritten Assamese strokes recognition. In: 2015 Twenty first national conference on communications (NCC). IEEE, pp 1–6
- Matsumoto K, Fukushima T, Nakagawa M (2001) Collection and analysis of on-line handwritten Japanese character patterns. In: Sixth International conference on document analysis and recognition, 2001. Proceedings. IEEE, pp 496–500
- Mehrotra K, Jetley S, Deshmukh A, Belhe S (2013) Unconstrained handwritten Devanagari character recognition using convolutional neural networks. In: Proceedings of the 4th international workshop on multilingual OCR. ACM, p 15
- Messaoud IB, Amiri H, El Abed H, Märgner V (2012) Region based local binarization approach for handwritten ancient documents. In: 2012 International conference on frontiers in handwriting recognition (ICFHR). IEEE, pp 633–638
- Mohiuddin S, Bhattacharya U, Parui SK (2011) Unconstrained Bangla online handwriting recognition based on mlp and svm. In: Proceedings of the 2011 joint workshop on multilingual OCR and analytics for noisy unstructured text data. ACM, p 16
- Mondal T, Bhattacharya U, Parui SK, Das K, Roy V (2009) Database generation and recognition of online handwritten Bangla characters. In: Proceedings of the international workshop on multilingual OCR, pp 1–6
- Mondal T, Bhattacharya U, Parui SK, Das K, Mandalapu D (2010) On-line handwriting recognition of Indian scripts-the first benchmark. In: 2010 International conference on frontiers in handwriting recognition (ICFHR). IEEE, pp 200–205
- Mukherjee S, Kumar P, Roy PP (2019) Fusion of spatio-temporal information for indic word recognition combining online and offline text data. *ACM Trans Asian Low Resour Lang Inf Process (TALLIP)* 19(2):1–24
- Murthy VN, Ramakrishnan AG (2011) Choice of classifiers in hierarchical recognition of online handwritten Kannada and Tamil aksharas. *J UCS* 17(1):94–106
- Namboodiri AM, Jain AK (2004) Online handwritten script recognition. *IEEE Trans Pattern Anal Mach Intell* 26(1):124–130
- Namboodiri A Team (2013) Development of OHWR system for Malayalam. *VishwaBharat J Lang Technol* 39&40:97–108
- Nethravathi B, Archana C, Shashikiran K, Ramakrishnan AG, Kumar V (2010) Creation of a huge annotated database for Tamil and Kannada ohr. In: 2010 International conference on frontiers in handwriting recognition (ICFHR). IEEE, pp 415–420
- Pal U, Chaudhuri B (1997) Automatic separation of words in multi-lingual multi-script Indian documents. In: Proceedings of the fourth international conference on document analysis and recognition, vol 2. IEEE, pp 576–579
- Pal U, Chaudhuri B (2003) Script line separation from indian multi-script documents. *IETE J Res* 49(1):3–11
- Parui SK, Guin K, Bhattacharya U, Chaudhuri B B (2008) Online handwritten bangla character recognition using hmm. In: 19th International conference on pattern recognition, 2008. ICPR 2008. IEEE, pp 1–4
- Plamondon R, Srihari SN (2000) Online and off-line handwriting recognition: a comprehensive survey. *IEEE Trans Pattern Anal Mach Intell* 22(1):63–84

- Prasad MM, Sukumar M, Ramakrishnan A (2009) Divide and conquer technique in online handwritten Kannada character recognition. In: Proceedings of the international workshop on multilingual OCR. ACM, p 11
- Prasanna SM, Choudhury H, Mandal S, Devnath S, Sundaram S (2013) Development of OHWR system for Gurmukhi script. *VishwaBharat J Lang Technol* 39&40:109–192
- Prasanth L, Babu V, Sharma R, Rao G, Dinesh M (2007) Elastic matching of online handwritten Tamil and Telugu scripts using local features. In: Ninth international conference on document analysis and recognition, 2007. ICDAR 2007, vol 2. IEEE, pp 1028–1032
- Primekumar K, Idiculla SM (2011) On-line Malayalam handwritten character recognition using wavelet transform and sfam. In: 2011 3rd International conference on electronics computer technology (ICECT), vol 1. IEEE, pp 49–53
- Rajkumar J, Mariraja K, Kanakapriya K, Nishanthini S, Chakravarthy V (2012) Two schemas for online character recognition of Telugu script based on support vector machines. In: 2012 International conference on frontiers in handwriting recognition (ICFHR). IEEE, pp 565–570
- Ramakrishnan A, Shashidhar J (2013) Development of OHWR system for Kannada. *VishwaBharat J Lang Technol* 39&40:67–96
- Ramakrishnan A, Urala KB (2013) Global and local features for recognition of online handwritten numerals and Tamil characters. In: Proceedings of the 4th international workshop on multilingual OCR. ACM, p 16
- Rampalli R, Ramakrishnan AG (2011) Fusion of complementary online and offline strategies for recognition of handwritten Kannada characters. *J Univ Comput Sci* 17(1):81–93
- Razavi S, Kabir E (2004) A data base for online Persian handwritten recognition. In: 6th Conference on intelligent systems, pp 859–863
- Reddy GS, Sarma B, Naik RK, Prasanna S, Mahanta C (2012a) Assamese online handwritten digit recognition system using hidden Markov models. In: Proceeding of the workshop on document analysis and recognition. ACM, pp 108–113
- Reddy GS, Sharma P, Prasanna S, Mahanta C, Sharma L (2012b) Combined online and offline Assamese handwritten numeral recognizer. In: 2012 National conference on communications (NCC). IEEE, pp 1–5
- Samanta O, Bhattacharya U, Parui SK (2014) Smoothing of hmm parameters for efficient recognition of online handwriting. *Pattern Recognit* 47(11):3614–3629
- Sanguansat P, Asdornwiset W, Jitapunkul S (2004) Online Thai handwritten character recognition using hidden Markov models and support vector machines. In: IEEE international symposium on communications and information technology, 2004. ISCIT 2004, vol 1. IEEE, pp 492–497
- Sharma A, Kumar R, Sharma R (2008) Online handwritten Gurmukhi character recognition using elastic matching. In: Image and signal processing, 2008. CISP'08. Congress on, vol 2. IEEE, pp 391–396
- Sharma A, Sharma R, Kumar R (2009a) Online handwritten Gurmukhi strokes preprocessing. *Int J Mach Graph Vis* 18:105–120
- Sharma A, Kumar R, Sharma R (2009b) Rearrangement of recognized strokes in online handwritten Gurmukhi words recognition. In: 10th International conference on document analysis and recognition, 2009. ICDAR'09. IEEE, pp 1241–1245
- Sharma A, Kumar R, Sharma R (2009c) Recognizing online handwritten Gurmukhi characters using comparison of small line segments. *Int J Comput Theory Eng* 1(2):131–135
- Sharma R, Sharma P, Sharma N, Singh H, Verma K, Kumar R, Kumar R (2013) Development of OHWR system for Gurmukhi script. *VishwaBharat J Lang Technol* 39&40:55–60
- Singh G, Sachan M (2015) A framework of online handwritten Gurmukhi script recognition. *Int J Comput Sci Technol (IJCST)* 6:52–56
- Singh S, Sharma A, Chhabra I (2016) Online handwritten Gurmukhi strokes dataset based on minimal set of words. *ACM Trans Asian Low Resour Lang Inf Process (TALLIP)* 16(1):1
- Singh H, Sharma R, Singh V (2018a) Efficient zone identification approach for the recognition of online handwritten Gurmukhi script. *Neural Comput Appl* 31:1–12
- Singh H, Sharma R, Singh V (2018b) Recognition of online unconstrained handwritten Gurmukhi characters based on finite state automata. *Sādhanā* 43(11):192
- Singh H, Sharma R, Kumar R, Verma K, Kumar R, Kumar M (2019) A benchmark dataset of online handwritten Gurmukhi script words and numerals. In: International conference on computer vision and image processing. Springer, pp 457–466
- Sundaram S, Ramakrishnan A (2008) Two dimensional principal component analysis for online character recognition. In: Proceedings of 11th international conference on frontiers in handwriting recognition (ICFHR), pp 88–94

- Sundaram S, Ramakrishnan A G (2009) An improved online Tamil character recognition engine using post-processing methods. In: 10th International conference on document analysis and recognition, 2009. ICDAR'09. IEEE, pp 1216–1220
- Sundaram S, Ramakrishnan A (2013) Attention-feedback based robust segmentation of online handwritten isolated tamil words. *ACM Trans Asian Lang Inf Process (TALIP)* 12(1):4
- Sundaram S, Ramakrishnan A (2015) Bigram language models and reevaluation strategy for improved recognition of online handwritten Tamil words. *ACM Trans Asian Low Resour Lang Inf Process* 14(2):8
- Swethalakshmi H, Jayaraman A, Chakravarthy VS, Sekhar CC (2006) Online handwritten character recognition of devanagari and Telugu characters using support vector machines. In: Tenth international workshop on frontiers in handwriting recognition. Suvisoft
- Takahashi K, Yasuda H, Matsumoto T (1997) A fast hmm algorithm for on-line handwritten character recognition. In: Proceedings of the fourth international conference on document analysis and recognition, 1997, vol 1. IEEE, pp 369–375
- Tan GX, Viard-Gaudin C, Kot AC (2009) Information retrieval model for online handwritten script identification. In: 2009 10th International conference on document analysis and recognition. IEEE, pp 336–340
- Tappert CC, Suen CY, Wakahara T (1990) The state of the art in online handwriting recognition. *IEEE Trans Pattern Anal Mach Intell* 12(8):787–808
- Veltman SR, Prasad R (1994) Hidden markov models applied to on-line handwritten isolated character recognition. *IEEE Trans Image Processing* 3(3):314–318
- Verma K, Sharma R (2015) Performance analysis of zone based features for online handwritten Gurmukhi script recognition using support vector machine. In: Progress in systems engineering. Springer, pp 747–753
- Verma K, Sharma RK (2016) Comparison of HMM-and SVM-based stroke classifiers for Gurmukhi script. *Neural Comput Appl* 28(1):51–63
- Verma K, Sharma R (2017a) An efficient writing-zone identification technique for online handwritten Gurmukhi character recognition. *Proc Natl Acad Sci India Sect A Phys Sci* 88:1–11
- Verma K, Sharma R (2017b) Recognition of online handwritten Gurmukhi characters based on zone and stroke identification. *Sādhanā* 42(5):701–712
- Wing AM (1979) Variability of handwritten characters. *Vis Lang* 13(3):283
- Zhang X-Y, Bengio Y, Liu C-L (2017) Online and offline handwritten Chinese character recognition: a comprehensive study and new benchmark. *Pattern Recognit* 61:348–360
- Zhou X-D, Yu J-L, Liu C-L, Nagasaki T, Marukawa K (2007) Online handwritten Japanese character string recognition incorporating geometric context. In: Ninth International conference on document analysis and recognition, 2007. ICDAR 2007, vol 1. IEEE, pp 48–52

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