CHAPTER-2
REVIEW OF LITERATURE

Gender classification and writer identification systems based on the handwritten text are useful applications that have significant participation in forensic analysis, criminal investigations, autopsy determination, suspected areas, crime analysis, verifying bank transactions etc. This chapter presents a series of deep surveys of the literature work on handwriting-based gender classification and writer identification systems. The detailed, comprehensive, and systematic review findings on both the systems have been collected from the reputed SCI Indexed journals, like expert systems with applications, pattern recognition, neural networks, artificial intelligence, and review, etc. with high impact factor. The proceedings of the International and national workshops and conferences such as International Conference on Document Analysis and Recognition (ICDAR), International Conference on Pattern Recognition (ICPR), IEEE conferences, etc. have been thoroughly considered for analyzing the research findings and studies. In this chapter, section 2.1 presents the state-of-the-art work on gender classification systems covering only non-Indic scripts because for Indic scripts the system has not been recognized so far. Section 2.2 presents the survey findings on the development of the writer identification system firstly with non-Indic scripts followed by Indic Scripts and then multi-scripts. Section 2.3 focuses on the research gaps and issues that have not been tapped by the researchers. So, to find the knowledge gap and unexplored area is the main objective of this chapter.

2.1 STATE-OF-THE-ART WORK ON HANDWRITING BASED GENDER CLASSIFICATION SYSTEM

Maken and Gupta (2021) presented a novel method for a gender classification system based on handwritten text using slants, area, and perimeter with K-NN, SVM, and logarithmic regression and achieved interesting results. Rahmanian *et al.* (2021) proposed the development of the gender classification system based on multi-script using the CNN approach. Cordasco *et al.* (2020) presented a gender classification system based on online handwriting using ANOVA analysis and achieved great success. Gattal *et al.* (2020) worked on the development of the gender classification system on QUWI datasets using the SVM approach. Illouz *et al.* (2018) presented a gender classification system with 405 participants in Hebrew and English scripts using CNN. Moetesum *et al.* (2018) developed a gender classification system on Arabic and English handwriting using convolution neural network (CNN) and linear discriminant analysis (LDA) classifiers and achieved a 70.08% accuracy rate. Morera *et al.* (2018) submitted a gender classification system for English and Arabic scripts using CNN. Nader *et al.* (2018) presented a novel model for age, gender, and nationality prediction based on handwriting characteristics and implementing geometric features, chain code features, edge features, and directional features with a 68.3% accuracy rate. Ahmed *et al.* (2017) proposed a gender classification system based on textural descriptors and hybrid classifiers with the majority voting, bagging, stacking approaches, Local Binary Patterns (LBP), Gray-Level Co-occurrence Matrices (GLCM), Histogram of Oriented Gradients (HOG), and Segmentation-based Fractal Texture Analysis (SFTA) with Nearest Neighbor (k-NN), Artificial Neural Networks (ANN), Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT) and achieved improved results on the QUWI database with 1017 writers. Akbari *et al.* (2017) constructed finite-state automata that help in generating feature vectors on QUWI datasets and MSHD datasets, with SVM and NN classifiers and got an accuracy rate of 79.3% and 77.8% with NN and SVM networks respectively on samples of QUWI datasets and accuracy rate of 79.9% and 79.0% with SVM and NN on samples of MSHD datasets.

Sahu *et al.* (2017) presented a gender classification through handwriting using the z-test and feature extraction method on 130 samples of males and females. Topaloglu and Ekmekci (2017) developed a framework for gender classification based on handwriting in Turkish script with 80 participants and tested with ID3 and J48 decision tree algorithms realizing a 93.78% accuracy rate. Upadhyay (2017) presented a deep survey on handwriting-based gender classification. Mirza *et al.* (2016) proposed a gender classification system based on the QUWI datasets and got promising results. Nogueras *et al.* (2016) presented a framework for text-dependent gender classification based on online uppercase handwriting on the BiosecurID database using pen-up strokes, attaining an accuracy of 76.0%. Tan *et al.* (2016) used transformational and geometrical features on ICDAR 2013 and RDF datasets and got an accuracy of 67.2% for gender classification. Bouadjenek *et al.* (2015a) developed a gender classification system using Local Binary Pattern (LBP) and Histograms of

Oriented Gradients (HOG) on the samples taken from the IAM dataset and realizing 70.0% accuracy. Bradley (2015) presented a survey on gender classification through handwriting samples. Kedar *et al.* (2015) proposed a personality identification system through handwriting and achieved a 90% accuracy rate. Levi and Hassner (2015) generated a model for age and gender classification systems using CNN and achieved high accuracy results. Siddiqi *et al.* (2015) developed a framework for the gender prediction on QUWI and MSHD datasets using Artificial Neural Network (ANN) and Support Vector Machine (SVM) classifiers on the features extracted from orientation, curvature, text-based features, and fractional features, thereby revealing 68.75% on QUWI and 73.02% with MSHD dataset. Youssef *et al.* (2013) worked on two datasets i.e., Arabic and English, for gender identification with 282 persons. Direction, curvature, chain code, and gradient direction features were exploited SVM classification technique and achieved an accuracy rate of 68.6% for Arabic and 85.7% in English script.

Table 2.1 illustrates the summarized results for the gender classification system. So, after having deep studies on the reviews, it is here concluded that the gender classification system based on handwriting for an Indic script has not been recognized yet and all findings presented here are non-Indic scripts either with a single script or with multi-script. To find an efficient framework and implementation for a gender classification system based on the offline handwritten text on the Gurumukhi script is a novel achievement and a challenging and useful task to achieve.

Table 2.1. State-of-the-art-work on Gender Classification System

Author/s	**Year**	**Single Script/ Multi-Script**	**Script**	**Feature Extraction and Classification Techniques**	**Accuracy Achieved**
Maken and Gupta	2021	Single	Roman	Slants, Perimeter, Area, Logistic Regression, SVM, K-NN, Majority Voting	High and satisfactory performance
Rahminian *et al.*	2021	Multi-script	IAM and KHATT	CNN	84.0% for GC and 99.15% for handedness
Cordasco *et al.*	2020	Single	Online handwriting	Time, Space, ANOVA analysis	Promising Results

Gattal *et al.*	2020	Single	QUWI dataset (Non-Indic script)	COLD and Hinge feature SVM classifiers	64.40%
Bi *et al.*	2018	Multi-script	Arabic, English, Chinese (Non-Indic script)	Character shapes, chain code, slant and curvature, SVM classifier	66.70%
Morera *et al.*	2018	Multi-Script	English, Arabic (Non-Indic script)	Deep CNN	80.72% with IAM datasets 68.90% with KHATT dataset.
Ahmed *et al.*	2017	Multi-Script	QUWI, Arabic, English (Non-Indic script)	Local Binary Patterns (LBP), Segmentation-based fractal texture analysis (SFTA). Gray-Level Co-occurrence Matrices (GLCM), Histogram of Oriented Gradients (HOG), SVM and ANN	85.0%
Akbari *et al.*	2017	Multi-Script	QUWI, MSHD (Non-Indic script)	SVM and NN	74.3% for Arabic 67.90% for English and French
Guerbai *et al.*	2017	Single	IAM (Non-Indic script)	Curvelet Transform, One class-SVM	77.33%
Mirza *et al.*	2016	Single	QUWI database (Non-Indic script)	Textural and Gabor Filter, Feed Forward Neural Network	70%
Tan *et al.*	2016	Multi-Script	ICDAR 2013 and RDF (Non-Indic script)	Geometrical and transformational features	67.2%
Bartle and Zheng	2015	Multi-Script	Blogs, formal handwritings (Non-Indic script)	WRCNN	86%
Bouadjenek *et al.*	2015a	Single	IAM dataset (Non-Indic script)	LBP, HOG	74.0%

Siddiqi *et al.*	2015	Multi Script	QUWI and MSHD (Non-Indic script)	Orientation, Curvature, text based, Fractional features. ANN&SVM	68.75% of QUWI database and 73.02% with MSHD datasets
Maadeed and Hussaine	2014	Multi Script	Arabic and English (Non-Indic script)	Random Forest and Kernel Discriminant Analysis	74.05% and 73.0% for same and different handwritten text
Youssef *et al.*	2013	Multi Script	Arabic and English (Non-Indic script)	Directional Features, SVM	68.6% for Arabic and 85.7% in English
Liwicki *et al.*	2011	Multi Script	Online and offline data (Non-Indic script)	Gaussian Mixture Model (GMM)	65.57%

2.2 STATE-OF-THE-ART WORK ON THE WRITER IDENTIFICATION SYSTEM

This section presented survey findings on the writer identification system in the handwritten text in Indic and non-Indic scripts both. In the first sub-section 2.2.1, survey findings have been considered based on the non-Indic and then summarized in Table 2.1 focusing on the parameters such as authors, year, script, feature extraction, and classification techniques and accuracy achieved. Table 2.1 also presents the results of developments of writer identification on multi-script too. In section 2.2.2, state-of-the-art work on the writer identification system based on the handwritten text in Indic script is presented with the tabular form in Table 2.2.

2.2.1 Based on non-Indic Scripts

Abbas *et al.* (2021) presented a textural measures-based writer identification system using local binary pattern and support vector machine on the single and multi-script non-Indic systems. Chen *et al.* (2021) presented an online writer identification system based on the variance of writing styles on the letter level and achieved great success. Hossain *et al.* (2021) put forward multi-zone character segmentation and merging approach with a convolutional neural network (CNN) deep model and attained 84%

precision for character and 82% precision for word level. Litifu *et al.* (2021) designed an offline writer identification system based on the redundant handwriting pattern by implementing dual-factor-analysis of variance (DF-ANOVA), Wigner distribution function and diagonal Index histogram, Fisher discriminant ratio (FDR) for feature selection and achieved 96.92% on IAM datasets and 96.4% on Firemaker dataset. Purohit and Panwar (2021) presented various text-independent writer identification methodologies based on the deep learning model. Wang *et al.* (2021) proposed a deep learning model on ICDAR17 datasets on historical document identification based on U-Net for digitization and Res-Net 50 for extraction of features.

Bensefia and Dieddi (2020) presented a double feature selection process and Fourier Elliptic transform based on graphemes on 100 writers of the IAM dataset and reported an identification rate of 96.0%. He and Schomaker (2020) proposed a new framework for writer identification based on the word or text block images, named FragNet using four benchmark datasets and reported good results. Javidi and Jampour (2020) worked for text-independent writer identification based on four non-Indic datasets, IAM, CERUG, FIREMAKER, and CVL, and achieved promising results. Sharma and Chanderiya (2020) presented a hand stroke and grapheme-based writer identification system using hand pressures by exploiting discrete cosine transform (DCT), principal component analysis (PCA), and support vector machine (S-SVM) were implemented and achieved promising 99.9% results. Patil and Mathur (2020) reported a comparison of numerous machine learning algorithms used for personality analysis and writer identification. Bennour *et al.* (2019) developed a model for writer recognition based on implicit shape codebook on CVL and BFL dataset and achieved great success. Chahi *et al.* (2019) presented an identification system on CVL, IFN/ ENIT, and IAM extracted LBP, LTP, and LPQ to achieve successful results. Chen *et al.* (2019) proposed an improvement in the performance of writer identification using a semi-supervised feature learning method named weighted label smoothing regularization. Lai and Jin (2019) developed a novel set of features for offline writer identification based on the path signature approach. A codebook method based on the log path signature showed competitive results on IAM, Firemaker, CVL, and ICDAR 2013 datasets.

Gattal *et al.* (2018) developed a method for gender classification on online multi-script handwriting images from QUWI dataset using oriented basic image and features (OBIF) and textural information such as local binary pattern, histogram of oriented gradients and Gabor filter are captured by implementing Support Vector Machine and retrieved max accuracy of 78.0%. Hadjadji *et al.* (2018) presented writer identification based on handwritten fragments, a clustered-based One-Class Classifier (OCC), Dynamic Fragment Weighting Combination (DFWC) rule to reduce the effect of inconsistent test fragments and achieved 97.56% and 94.51% on IFN/ENIT and IAM datasets respectively. Halder *et al.* (2018) proposed local handwriting-based attributes, multi-layer perceptron, conventional segmentation based methods, and simple logistic classifiers on a handwritten Bangla script with 190 writers. Rehman *et al.* (2018) developed a comprehensive review on the writer identification system and methodologies by providing a taxonomy of dataset, feature extraction methods, as well as conventional and deep learning classification models for writer identification for English, Arabic, and western scripts. Chahi *et al.* (2018) worked on the same datasets with LBP thereby attaining satisfying results. Kallel *et al.* (2018) proposed a writer identification system for Arabic handwriting using the curvelet transform method for the feature extraction process, with an SVM classifier by collecting samples of the KHATT dataset from 100 authors. In the feature extraction process with multi-resolution levels, the experiment presented superiority and high identification rates. Ahmed *et al.* (2017) presented a writer identification and gender classification system based on textural features with ANN, SVM, discrete transform, and random forest using bagging, voting, and stacking methods on QUWI datasets. Ashiquzzaman and Tushar (2017) proposed a method for identification of handwritten numerals in Arabic script using multilayer perceptron method (MLP) and convolution neural network (CNN), thereby achieving an accuracy of 97.4%. Asi *et al.* (2017) developed a novel approach for writer identification using Arabic historical manuscripts collected from WAHD and KHATT datasets. They converted contour direction features into global features, and classify using the nearest neighbor approach. Durou *et al.* (2017) proposed a combination of Oriented Basic Image features, graphemes codebook, and kernel PCA to reduce the high dimensionality of feature values on the samples collected from the IAM dataset for English handwriting and ICFHR 2012 dataset for Arabic handwriting and achieved 96% accuracy rate.

Fukue *et al.* (2017) presented a framework for writer identification system on the samples collected in Thai script using Individual Change Control Processing, ICCP and received an improved accuracy rate from 98.44% to 99.54% of Thai characters. Ghosh and Maghari (2017) put forward a method for the recognition of handwritten digits using Deep Neural Network (DNN), Convolutional Neural Network (CNN), and Deep Belief Network (DBN) and achieved an accuracy rate of 98.08% with DNN. He and Schomaker (2017) proposed two novel curvature-free features, namely, run length of local binary patterns (LBPruns) and Cloud of line distribution (COLD) that worked on the samples of CERUG, IAM, and Firemaker datasets. Tan *et al.* (2017) presented a systematic and incredible review on state-of-the-art work on offline text-independent writer identification based on three scripts named, English, Chinese, and Arabic scripts and extracting texture, allograph, and structure features. Xiong *et al.* (2017) proposed public datasets for writer identification and compared the performance of existing methods based on frequency domain features and realized that spatial distribution features are superior to both frequency domain features and shape features in capturing the individual traits. Mohsen *et al.* (2017) proposed a novel technique for the identification of writers based on deep learning and used a stack de-noising autoencoder for extracting features with a support vector machine for classification. Alwzwazy *et al.* (2016) generated a deep learning-based convolutional neural network (CNN) for the handwritten Arabic script with 45000 samples and achieved an accuracy of 95.7%. Dhieb *et al.* (2016) proposed a writer identification system based on a deep neural network (DNN) and beta elliptic model for online data by considering the writing movements of the writer using profile entities. Khan *et al.* (2016) presented writer identification in Arabic and English script using multiscale ternary pattern and discriminant analysis by implementing the majority voting scheme and achieved 99.4% on IAM and 87.5% on AHTID datasets.

Maadeed *et al.* (2016) worked on the text-dependent writer identification on the Arabic handwriting of 100 writers and achieved a 90.0% accuracy rate with edge detection probability distribution and K-Nearest Neighbor classifiers. Poznanski and Wolf (2016) developed a method for the images of handwritten words with multiple fully connected branches and CNN to evaluate its n-gram frequency profile. Frequencies for unigrams, bigrams, and trigrams were estimated for the entire word. Canonical correlation analysis was used for the comparison of the profiles of all words and high accuracy was achieved. Abdi and Khemakhem (2015) presented a grapheme-based methodology for identifying and verifying writers in Arabic script by taking samples from 411 writers of the IFN/ ENIT dataset and 60 feature vectors. Newell and Griffin (2014) developed writer identification in Roman script with 300 writers using oriented basic image features and with Euclidean distance, thereby achieving 99.0% accuracy. Wu *et al.* (2014) presented a system for identifying writers, including training, enrollment, and identification, and implemented it on six data sets. Brink *et al.* (2012) presented ink width pattern-based writer identification based on Quill and Quill Hinge features and achieved an accuracy rate of 70 to 90%. Akbari *et al.* (2012) worked on the Persian script on 50 writers using wavelet transformations & co-occurrence matrix. They used a K-NN classifier and achieved an accuracy of 93.3% for the writer identification. Siddiqi and Vincent (2010) proposed a writer identification system based on the orientation and curvature patterns and graphemes and achieved an accuracy rate of 96.2%.

Rahiman and Rajasree (2009) carried out a novel approach for recognizing printed character recognition using the back-propagation neural network, using wavelet transform features, wavelet multi-resolution analysis, and feed-forward backpropagation NN on 715 images and achieved an accuracy of 92.0%. Bulacu and Schomaker (2007) successfully developed a writer identification system for the Arabic handwriting with the IFN/ENIT datasets on 350 writers using textural features, allographic features, and Nearest Neighbor classifier, achieving promising accuracy of 99.0%. Dmour and Zitar (2007) exploited hybrid spectral and statistical measures (SSMs) of texture for Arabic writer identification and compared them with multiple-channel Gabor filters and the grey-level co-occurrence matrix (GLCM) by exploiting four classifiers namely, Linear Discriminant Classifier (LDC), Support Vector Machine (SVM), Weighted Euclidean Distance (WED), and the K-Nearest Neighbor (K-NN) classifier, achieving 90.0% accuracy. Nejad and Rehmati (2007) implemented multi-channel Gabor filters and weighted Euclidean distance in developing a writer identification system for 40 writers and achieved 80.0% accuracy. Siddiqi and Vincent (2007) presented a writer identification system on a roman script using samples from IAM datasets of 50 writers and extracted features on the basis of window positioning and Bayesian classification method, thereby achieving 94.0%.

2.2.2 Based on Indic Scripts

Sharma and Kaushik (2020) presented a comprehensive and systematic survey in recognition of handwritten characters in Indic scripts and proposed a framework based on CNN and swarm optimization. Girdher and Sharma (2020) explored a comprehensive survey on the writer identification system for Indic scripts mainly Bengali, Devanagari, Gujarati, Gurumukhi, Kannada, Malayalam, Oriya, Tamil, Telugu, and found that the developments on the Indic scripts are limited as compared to the non-Indic scripts due to non-availability of the dataset. Mukherjee and Ghosh (2020) proposed a framework based on the genetic as well as memetic factors on the English handwritten text samples of Bengali script using stroke features like a doughnut, hump, stick, and stem-loop, MLP, and K-STAR classification methods and achieved 93.54% and 95.69% respectively. Adak *et al.* (2019) performed writer identification based on offline handwriting in Bangla script using handcrafted features and support vector machine and achieved promising results. Ahmed *et al.* (2019) presented a deep multidimensional network and CNN was employed with Urdu Nastaliq Handwritten Dataset (UNHD) samples to recognize Urdu handwriting and achieved good accuracy. Garg *et al.* (2019) worked on the character recognition in the Gurumukhi script using PCA, linear, polynomial, and RBF SVM and k-NN on 160 writers and achieved a 92.3% accuracy rate.

Sakshi *et al.* (2018) presented a writer identification system on the Gurumukhi script with zoning, diagonal, transition, intersection and open endpoints, centroid, horizontal peak extent, vertical peak extent, parabola curve fitting, and power curve fitting-based features with Naive Bayes, Decision Tree, Random Forest, and AdaBoostM1 classification techniques on 49,000 samples collected from 70 different writers and revealed maximum accuracy of 81.75% with centroid features and AdaBoostM1 classifier. Prabhanjan and Dinesh (2017) proposed a unique approach for the recognition of the Devanagari script using a deep belief network and the unsupervised restricted Boltzmann machine and achieved an accuracy of 83.44% with unsupervised method and accuracy of 91.81% with the supervised method. Verma and Sharma (2017) developed a novel zone identification method of character recognition of the online handwritten 4280 Gurumukhi characters from 10 writers and using a hidden Markov model (HMM) for 74 different stroke classes with 5-fold cross-validation and produced an accuracy rate of 95.3%. Kalra and Rani (2017) developed a writer identification system with 30 writers in the Gurumukhi script based on open and endpoint intersection, zoning features with MLP classifier for classification, and attained an accuracy rate of 53.0%. By applying majority voting-based rejection criteria, 97.0% identification accuracy was achieved.

Aggarwal and Sharma (2016) exploited a discrete Fourier transform and RBF kernel for identifying online handwritten Gurumukhi text using 8408 stroke samples and attained 91.7% accuracy. Kumar *et al.* (2016) put forward a system for offline handwriting grading for the writers with zoning, Zernike moments, open-end, diagonal, and intersection features and for classification hidden Markov Model (HMM) and Bayesian classifiers were implemented on the Gurumukhi characters. Desai (2015) proposed a writer identification system with a multi-layered feed-forward neural network for the classification using thinning and skew correction in the Pre-processing phase, thereby achieving 82.0% accuracy. The network was implemented on the trained data of 610 digits and tested data of 2650 digits. Dhandra *et al.* (2015) implemented a text-dependent writer identification method based on Kannada handwriting samples from 25 writers using directional multi-resolution spatial features, Radon Transform, structural features, and Discrete Cosine Transform and attained 93.25% using 5-fold cross-validation. Halder *et al.* (2015) presented a method of Devanagari script on 50 writers, 5 copies of handwritten characters from each writer, and using LIBLINEAR and LIBSVM classifiers of WEKA environment to get the writer of the characters, achieving promising results of 99.12% with LIBLINEAR and all writers. Sagar and Pandey (2015) generated a system for identifying Devanagari/Hindi documents using ANN and exploited slant estimation techniques, Hough transforms and Zernike moments.

Purkait *et al.* (2010) developed a writer identification system based on the handwritten Telugu documents using directional morphological features, including k-curvature, directional opening, directional erosion, directional curvature, and used Nearest Neighbor classifier with leave-one-out strategy, achieving a maximum accuracy of 82.70% for a word with word index 6 and achieved an accuracy of 71.73%, 86.89%, 93.12%, and 98.21% for one word, two words, three words, and four words, respectively with the directional opening method. Pitak and Matsuura (2004) worked for a writer identification system for Thai handwritten documents using 81 writers, and implemented velocities of the pen movements, Fourier transforms approach as a technique for feature extraction, and successfully got the identification accuracy of 98.5%. Soma *et al.* (2014) presented writer identification on 100 writers, 95.12%, offline kanji characters. Srreraj (2012) proposed a thesis on writer identification based on Malayalam script using different feature extraction and classification algorithms and achieved successful results. Srreraj and Indicula (2011) developed a writer identification system on a Malayalam script based on graphemes and neural networks on the samples collected from 280 writers and achieved an 89.28% accuracy rate. Biswas and Das (2012) presented a novel approach for extracting two different sets of components, namely, fragment Set-A and fragment Set-B, and tested for one script having 90-110 words and another script having 20-30 words separately and achieved 92.72% and 80.0%, respectively. Dhandra *et al.* (2012) used Gabor filtering, gray level co-occurrence matrix and discrete cosine transform for writer identification on 20 writers in Kannada scripts using the K-NN classification method and concluded Gabor energy features are more helpful in achieving accuracy as compared to DCT and GLCMs based features. Chanda *et al.* (2010) proposed a method for the identification of individuals based on the Bengali handwritten samples of 104 writers, using discrete directional and support vector machines, achieving a maximum identification rate of 99.03%. Hiremath *et al.* (2010) presented a novel approach for writer identification of handwritten documents in Kannada by implementing discrete wavelet transform and K-NN classifier with 30 writers and got promising accuracy of 91.45%. Majumdar (2007) proposed a character recognition system for Bangla script using digital curvelet transform, k-Nearest Neighbor, ridgelet transform and discrete ridgelet transform features for the extraction of features and received maximum accuracy of 97.35%.

2.2.3 Based on multi-scripts

BabaAli (2021) presented an online writer identification system on the samples of CAISA datasets in English and Chinese scripts, using class covariance and linear discriminant analysis, and achieved 98.68% and 96.03% respectively. Sabzekar *et al.* (2021) worked on the writer identification on the data samples in Persian and Roman script, using texture features and wavelet transform, and classify the data using MLP. Sheng *et al.* (2021) presented a novel framework on the writer identification system on four public datasets using global and local fragments-based features forming hierarchical attention pooling and global recurrent neural network. Mohammed and

Ahmed (2021) developed a writer identification system on the KRDOH and IAM datasets of Kurdish handwritten datasets with 1076 writers and achieved an accuracy rate of 94.3%. Hossain *et al.* (2021) proposed a multi-zone character segmentation and merging approach with a convolutional neural network (CNN) deep model and attained 84.0% precision for character and 82.0% precision for word level. Bhunia *et al.* (2020) presented Indic handwriting recognition based on offline and online data in English and six official Indic scripts and attained promising rates. Kumar and Sharma (2020) proposed segmentation and Pre-processing free SEG-WI model based on a deep convolution neural network (DCNN) for offline text-independent writer identification and achieved an identification rate of 92.79%, 99.35%, 98.30%, 87.06% for IAM, CVL, IFN/ENIT, and Devanagari, respectively. Kumar and Sharma (2019) explored a new idea based on the Distribution Descriptive Curve (DDC) and Cellular Automata (CA). DDC used pixel distribution text images for generating a unique curve for generating features and worked on four databases IAM, IFN/ENIT, and Indic scripts namely, Kannada and Devanagari (Hindi), and achieved promising results.

Christlein *et al.* (2018) performed a writer identification system on ICDAR17 datasets, RNN, and deep residual network on the training datasets, CNN, and achieved good results for identification. Shaikh *et al.* (2018) presented a hybrid deep learning model for writer identification. It is an amalgamation of Auto-Learned Features (ALF) and Human-Engineered Features (HEF), Autoencoder, CNN, gradient structure concavity, and scale-invariant feature transform and achieved the maximum of 99.17%. Aubin *et al.* (2017) proposed online writer identification based on the pen pressure through SVM and K-fold cross-validation techniques and achieved 95.0% accuracy. Adak *et al.* (2017) implemented a writer identification method with Bangla and English with 29,341 words with the handcrafted feature set and used a support vector machine with radial basis function (RBF) to achieve 79.07% identification accuracy. Christlein *et al.* (2017) proposed offline writer identification on ICDAR, CVL, and KHATT datasets using GMM and Root SIFT descriptors and achieved high accuracy rates. Kumar *et al.* (2017) presented a writer identification system using structural properties of the character for the feature extraction, using Devanagari and Latin datasets.

Roy *et al.* (2017) proposed an endeavor using deep belief networks for compressed delineation of sequential data, and the Hidden Markov Model (HMM) for word recognition on RIMES and IFN/ENIT datasets on Latin and Arabic scripts and also tested with Devanagari datasets. Yang *et al.* (2016) presented a novel approach that uses an end-to-end DeepWriterID which is based on the deep Convolutional Neural Network (CNN) method two data sets included were 187 writers for Chinese and second with 134 writers in English and achieved an identification rate of 95.72% for Chinese text and 98.51% for English text. Bouadjenek *et al.* (2015b) proposed a model for age, gender, and handedness prediction based on the histogram of oriented features, local binary pattern, distribution of gradient patterns on Arabic and Roman datasets using the SVM classifier. He *et al.* (2015) developed a writer identification system based on junction detection, using a CERUG dataset consisting of Chinese and English scripts. Khalid and Navqi (2015) presented a writer identification system in French and Arabic scripts by implementing codebook features on the standard LAMIS MSHD database for 87 writers and used Chi-Square, k-NN, and Euclidian test. Surintan *et al.* (2015) proposed a system for the handwritten character recognition by using local gradient feature descriptors and Histogram of Oriented Gradients (HOG) with K-NN and Support Vector Machine (SVM) methods on the data samples collected in three different scripts namely Thai, Bangla, and Latin and achieved promising rate. Dhandra *et al.* (2014) exploited a text and script independent method on 100 writers in Kannada, Roman, and Devanagari script using grey level co-occurrence matrix and achieved 82.19 % from Kannada - Roman, and Devanagari documents.

Thendral *et al.* (2015) presented RBF and Kernel-based model for classification of the writer in Tamil handwriting using SVM. Thendral *et al.* (2014) worked on loops, curves, directions of the character and using SVM performed writer identification in Tamil script and attained 90.06% using the RBF kernel. Bertolini *et al.* (2013) performed writer identification on the samples of the Brazilian forensic letter dataset and IAM dataset using linear binary pattern and linear phase quantization features and achieved the maximum 99.2% accuracy. Chan *et al.* (2008) presented a writer identification system based on a distance between the distributions of patterns at the character level on online handwriting using 82 samples and achieved 95%. Li *et al.* (2007) produced a model for writer identification with 242 writers.

They used hierarchical structure, fusion dynamic features, and static features. They preferred the k-NN classifier and got an accuracy of 90.0% for Chinese and 93.0% for English. Baghshah *et al.* (2006) presented an online Persian handwriting model using fuzzy learning vector quantization and achieved high accuracy. Bensefia *et al.* (2005) generated a writer identification system based on the textual based information retrieval model and building feature space using PSI and IAM dataset and achieved the maximum 96.0% accuracy rate. Tomai *et al.* (2004) presented gradient, structural, and concavity-based features and worked on word, shape curvature, and shape context to achieve a satisfactory accuracy rate.

Table 2.2 and Table 2.3 represent the work done on the Writer Identification System in non-Indic and Indic scripts, respectively.

Table 2.2. State-of-the-art on Writer Identification System based on Non-Indic Scripts

Author	Year	Single script/ Multi script	Non-Indic Script	Feature Extraction and Classification Techniques	Accuracy Rates
BabaAli	2021	Multiscript	English and Chinese	class covariance and linear discriminant analysis (LDA)	98.68% and 96.03%
Litifu *et al.*	2021	Single	IAM, FIREMAKER	DF-ANOVA, Dual factor- Analysis of Variance and Wigner distribution function, Diagonal Index histogram. Fisher discriminant ratio (FDR).	96.92% on IAM, 96.4% on Firemaker.
Abbas *et al.*	2021	Single and Multiscript	BFL, KHATT, MSHD, CERUG, WDAD	Textural measures, linear binary pattern, Oriented Basic Image Features, SVM	90.5%
Sharma and Chanderiya	2020	Single	Roman (22000 grapheme samples)	Handstroke and Grapheme based features Discrete-Cosine Transform, Principal-Component-Analysis and Support Vector Machine (S-SVM)	99.9%

He and Schomaker	2020	Multiscript	IAM, CVL, Firemaker and CERUG-EN	Word or text blocks images Deep neural network, named FragNet,	95.0% to 99.9% with four datasets
Mukherjee and Ghosh	2020	Single	English handwritten text samples from Bengali origin and Bengali-medium schooling background	Genetic as well as Memetic factors.	MLP and K-STAR are 93.54% and 95.69% respectively.
Kumar and Sharma	2020	Multiscript	IAM, CVL, IFN/ENIT, Kannada, and Devanagari	segmentation and Pre-processing free SEG-WI model with deep convolution neural network (DCNN)	92.79%, 99.35%, 98.30%, 100.00%, 87.06%
Kumar and Sharma	2019	Multiscript	IAM, IFN/ENIT, Kannada and Devanagri	Descriptive Curve and Cellular Automata (DDC)	97.8% with IAM datasets
Kallel*et al.*	2018	Single	Arabic	Pyramid Decompositions, SVM classifier	87.5%
Chahi *et al.*	2018	Multiscript	Arabic and English IFN/ENIT, AHTID/MW, CVL and IAM	Block wise local binary count (BW-LBC), Euclidean, correlation, Bhattacharyya distance	AHTID/MW 99.53%, IFN/ENIT 97.56% IAM 90.11% CVL 99.03%
Adak *et al.*	2017	Multiscript	Bangla and English	SVM, RBF	79.07%
Chahi *et al.*	2017	Multiscript	Arabic and English	Histograms, Blockwise Local Binary Count	98.76%
Durou	2017	Multiscript	IAM dataset for roman and ICFHR dataset for Arabic	OBI features and Graphemes, K-NN	96.0%
Fukue	2017	Single	Set of Thai characters	Individual change control processing Euclidean distance	99.5%

He and Schomaker	2017	Multi	Firemaker and IAM	Run-lengths of Local Binary Pattern (LBPruns) and Cloud of Line Distribution (COLD), Nearest Neighbor classifier with leave one out method	96.6% on Firemaker 96.9% on IAM
Alwzwazy *et al.*	2016	Single	Arabic	Deep Convolutional Neural Network	95.7%
Khan *et al.*	2016	Multi	IAM and AHTID/MW	Multiscale local ternary pattern & Spectral Regression Kernel Discriminant Analysis Histogram Predictor Models and majority voting	99.4% for IAM Datasets and 87.5% for AHTID/MW Database
Maadeed AI *et al.*	2016	Single	Arabic	Edge-based directional probability distributions K-Nearest Neighbor	90.0%
Hannand *et al.*	2016	Multiscript	IFN/ENIT and IAM dataset	Local Binary Patterns, Local Ternary Patterns and Local Phase Quantization Histograms using textural information	IFN/ENIT Arabic,94.89% IAM English, 89.54%
Zhu and Wang	2016	Multi script	Chinese and English	Sparse Auto Encoder Codebook, K-NN	98.6% with IAM database And99.2%for HIT-MW database
Soma *et al.*	2014	Single	Japanese	Local and Global features Majority voting scheme	99.0%
Chu and Srihari	2014	Single	Roman	Word level features, Deep Neural Network	Satisfactory results
Bertolini *et al.*	2013	Multiscript	Brazilian forensic letter database and IAM database	Texture based descriptor, SVM with 5-fold cross validation	96.7% with IAM,99.2% with BFL
Saranya and Vijaya	2013	Single	Roman	Edge direction distribution and edge hinge distribution Support Vector Machine	94.27% accuracy for word level and 90.10% for character level.

Akbari *et al.*	2012	Single	Persian	Wavelet Transform & Co-occurrence matrix K-NN	93.3%.
Awaida and Mahmoud	2012	Single	Arabic	Connected component, Gradient distribution Nearest Neighbor K-NN	95.4%
Ram and Moghaddam	2009	Single	Persian	Grapheme and Gradient, Fuzzy Classifier	90.0%
Nejad and Rahmati	2007	Single	Farsi	Multi-channel Gabor filtering with moments, the weighted Euclidean distance	80.0% at text level 100% on word level
Li *et al.*	2007	Multiscript	English and Chinese	Hierarchical Structure in Shape Primitives + Fusion Dynamic and Static Features Nearest Neighbor K-NN	90.0% for Chinese and 93.0% in English
Schomaker and Bulacu	2004	Single	Firemaker dataset	Connected-component contours Nearest Neighbor and Chi square Distance	97.0%
Zhang *et al.*	2003	Single	Latin	Gradient (192 bits), Structural (192 bits) and Concavity (128 bits), k-Nearest Neighbor classification	97.7%
Srihari *et al.*	2002	Single	Roman	Gradient, structural, and concavity histograms. Eleven macro features Euclidean distance, Correlation measure	98.0%

Table 2.3. State-of-the-art on Writer Identification System based on Indic Scripts

Author	**Year**	**Single script / Multi script**	**Indic Script**	**Feature Extraction and Classification Techniques**	**Accuracy Rates**
Dargan *et al.*	2020	Single	Devanagari	Zoning, Diagonal, Transition, Peak Extent, K-NN, SVM	91.53%
Kumar *et al.*	2018	Single	Gurumukhi	K-NN, SVM, Zoning, Diagonal, Transition, Peak Extent	89.85%
Sakshi *et al.*	2018	Single	Gurumukhi	Zoning, diagonal, transition, intersection and open-end points, centroid, Naive Bayes, Decision Tree, Random Forest and AdaBoostM1	81.75%
Prabhanjan and Dinesh	2017	Single	Devanagari	Unsupervised restricted Boltzmann machine with deep belief network	83.44% and 91.81% with unsupervised and supervised, respectively
Andrew *et al.*	2017	Single	Telugu	Descriptive convolution-based features using directional filters, NN,SVM	99.3%
Adak *et al.*	2017	Single	Bengali	Support Vector Machine with Radial Basis Function	79.1% Bengali
Kalra and Rani	2017	Single	Gurumukhi	MLP, Open end point, zoning features	53.0%
Hangarge *et al.*	2016	Single	Kannada	DCT, Gabor Filter, Gabor Energy and GLCM,K-NN	DCT 77.0%, Gabor energy 88.5%, GLCM 79.5%
Desai	2015	Single	Gujarati	Multilayer Feed Forward Neural Network	82.0%

Misra	2015	Single	Odia	Contour based features Back Propagation Neural Network and HMM	84.5%
Halder	2015	Single	Devanagari	LIBLINEAR, LIBSVM	99.12%
Adak and Chaudhari	2015	Bangla	Local Binary Pattern, Pseudo structural Descriptor, Index of Dissimilarity	Multiple SVM classifiers With RBF kernel	86.6%
Halder and Roy	2013	Bangla	400 dimensional	MLP	99.1%
Chanda *et al.*	2012	Single	Oriya	Directional chain code and Curvature feature, SVM	99.0%
Choksi and Thakkar	2012	Single	Gujarati	Structural, geometric and Wavelet, K-Nearest Neighbor, General Regression Neural Network, Fuzzy KNN	K-Nearest Neighbor 67%, General Regression Neural Network 97%, Fuzzy KNN 100%
Biswas and Das	2012	Single	Bangla	Distance matrix, Histogram	92.72%
Jayanthi and Rajalakshmi	2011	Single	Tamil	Texture analysis, gray level co-occurrence matrix, Maximum Correlation Coefficient	82.8%.
Hiremath *et al.*	2010	Single	Kannada	Discrete wavelet transforms, K-NN	91.5%
Desai	2015	Single	Gujarati	Horizontal, vertical and two diagonal features Multi Layered Feed Forward Neural Network	82.0%

Chanda *et al.*	2010	Single	Bengali	Directional and Gradient Support Vector Machine	99.0%
Prasad *et al.*	2009	Single	Kannada	Segmented stroke group PCA	81.0%
Roy and Paul	2009	Single	Bangla	Structural features, Deep Neural Network	95.7%

After having a thorough review of the developments of the writer identification system by using different datasets and different feature extraction and classification techniques, it has been concluded that work done with handwritten text in non-Indic is very much recognized with very high acceptable accuracy rates in comparison to Indic scripts. Also, the development of such a system in the Gurumukhi script is rarely recognized and acknowledged with less accuracy rate. So, the objective of development of the writer identification system in Gurumukhi script with a sufficiently large number of writers is really a novel endeavor.

2.3 RESEARCH GAPS AND CRITICAL ANALYSIS BASED ON STATE-OF-THE-ART WORK ON GENDER CLASSIFICATION AND WRITER IDENTIFICATION SYSTEM

- Owing to diverse handwriting styles, offline handwriting identification is undoubtedly a difficult task.
- It is perceived that the gender classification system for offline handwritten text in Gurumukhi script is a novel area to explore and a novel contribution for handwriting-based research in forensic investigations.
- Regarding the development of the writer identification system in the Gurumukhi script, limited work has been recognized on a small size dataset, so the proposed work is an exigent and demanding application with a large dataset and with an improved accuracy rate.
- Standardized datasets always play a significant role in the research process. As such, no benchmark corpus in Gurumukhi script has already been available with a sufficient number of writers along with their name and gender to perform writer identification and gender classification, so this is also an

original realization to generate a corpus for the development of proposed applications.

- In the survey findings, it has been noticed that the accuracy rate for writer identification in the Gurumukhi script can be improved further by maintaining a quality dataset and implementing efficient Pre-processing, feature extraction, and classification techniques.
- Hybridization of feature extraction techniques can be implemented so as to attain a successful identification accuracy rate.
- Hybridization of classification techniques can also be another approach for boosting the accuracy with Gurumukhi scripts to boost the accuracy rate.
- Dimensionality reduction is another good choice for extracting only significant feature values and eliminating the extraneous variables. This can be another perspective with gender classification and writer identification that has not been explored with Gurumukhi script.
- Also with Indic script, the gender classification system and writer identification systems can be developed based on the multi-script, including Indic and non-Indic scripts, for training and testing, and hence it generates new directions for handwriting-based researchers.
- Development of writer identification and gender classification with online handwriting in Gurumukhi script can be another area of research.
- Another gap that can be filled is to inculcate the integrated approach i.e. hybridization of feature extraction with hybridization of classification methods to attain high accuracy.
- The proposed work is a novel, original and challenging research work in concern with the gender classification system, i.e., to develop the framework for a gender classification system based on offline handwritten text in Gurumukhi script, an Indic script.
- To propose a framework for the writer identification system for Gurumukhi script with improved accuracy results and with a large dataset is a great realization in the field of document analysis and forensic investigations.

From the state-of-the-art work, it is perceived that work done for the writer identification and gender classification system in non-Indic scripts based on the handwritten text is very much recognized with acceptable accuracy rate. For Indic scripts, especially Gurumukhi script, the development of an efficient system for writer identification and gender classification based on large and qualitative dataset with efficient methodologies is the major and crucial issue for the researchers.

2.4 CHAPTER SUMMARY

This chapter presents deep and systematic and comprehensive literature survey on both the systems i.e., gender classification and writer identification. The objective of this chapter is to find the unexplored issues and experiences a deep insight into the concept, approaches, the nature of the dataset, classifiers, and accuracy rate achieved so as to acquire the motivation and novel ideas for the optimum developments. In this chapter, firstly state-of-the-art work has been presented on the gender classification system by covering both Indic and non-Indic scripts, beginning with the current year. Then we have a literature survey on the writer identification system covering Indic, non-Indic scripts, and multi-script. Research and knowledge gaps and critical issues have been discussed elaborately in section 2.3.