ISSN: 2752-3829 Vol. 3 No.2, (December, 2023)

Stochastic Modelling and Computational Sciences

A SYSTEMATIC ANALYSIS OF TECHNIQUES FOR IDENTIFYING WRITERS BASED ON HANDWRITING

SukhdeepKaur[1] and Ganesh Kumar Sethi[2]
[1]Research Scholar, Punjabi University Patiala Punjab, India-147002 and [2]Associate Professor, M.M. Modi College Patiala-147001
[1]skjassal91@gmail.com and [2]ganeshsethi147@gmail.com

ABSTRACT
This paper presents an overview of research conducted on writer identification algorithms, datasets, methodologies, and the organization of competitions. The primary goal of this research paper is to provide an overview of the work done one on non-Indic scripts like Chinese, Arabic, French, Roman and Persian as well as Indic scripts like Bengali, Gujarati, Gurumukhi, Kannada, Malayalam, Oriya, Tamil, and Telugu. The study also describes various phases of identification process. It aids the researchers in field of identification by lending the studies of different pre-processing methods, techniques for feature extraction and classification based on various Machine learning and Deep learning classifiers on both scripts- Indic and non-Indic. Compared to the scripts of non-Indic with a high accuracy rate, there has been comparatively less research conducted on writer identification in Indic scripts. The availability of dataset of Indic scripts also less. The study also incorporate an overview of diverse datasets availability along with source and competitions conducted within the domain.

Keywords: Handwriting, Identification, Writer, Classification, Machine learning, Deep learning.

1. INTRODUCTION
Today, there is a significant demand for the identification of a writer's handwriting, driven by the immense advancements in technology and its widespread applications across various fields. This method involves using a reference training database to compare a text to determine who wrote it. The complexity arises from the fact that every person's writing style is inherently unique, influenced by their emotions, perceptions, behaviors, and even their cognitive processes. The emergence and advancement of artificial intelligence as well as pattern recognition techniques have spotlighted the research challenges associated with identifying writers based on their handwriting. A subfield of computer science called biometric identification is concerned with separating individuals from a wider population based on a variety of identifiers, including fingerprints, retinal patterns, handwriting, and signatures. On comparing with non-Indic scripts, the efficiency of work carried out on the writer identification systems in Indic scripts is lower. Identifying writers on the basis of handwriting includes a number of challenges, for instance, extraction of specific features of hand-writing, representation of features by using suitable approaches, and developing various novel techniques for feature extraction and classification. It also requires a proper dataset with sufficient writers. There are several languages across the world. Every language has its writing style and characteristics. It states that the identification problem also varies for different languages.

2. APPLICATIONS OF WRITER IDENTIFICATION SYSTEM
The application areas of identifying writer involve forensic expert decision-making systems, historical documents analysis, validation, and authentication purposes, like identification and verification of signatures are crucial procedures employed both in banking institutions and within the judicial system.

- **Forensic Document Analysis:** Handwriting analysis is categorized within the domain of questioned documents section of forensic science and they are as unique as a fingerprint which focuses on analyzing various characteristics of handwriting and writing style to establish a connection between a known writer and an unknown document.
- **Criminal Justice System:** Written documents are utilized as evidence in the court of law for determining the guilt or innocence of those accused of the crimes against society.

- **Biometric Modality:** Biometric identification is a technology that employs scanned graphical data from multiple sources for the purpose of personal identification. This technology assists organizations in safeguarding their priceless collections, infrastructure, and human resources, enhancing safety and security measures.
- **Validation and Authentication as Signature Identification:** There is a need for adequate protection of signatures and there is need for a system, that can identify who is the signatory with a great degree of certainty.
- **Personal Identification:** Because handwriting is unique to each individual, it is used to verify that a document is written by that person, indicating the potential use of the writer identification system.
- **Mobile Bank Transactions:** These systems enable banks and financial institutions to streamline the automated extraction and categorization of data from various supporting documents, including loan applications, handwritten notes, envelopes, and more.
- **Authentication Access to Network:** In order to secure a variety of resources, such as corporate buildings or electronic assets, authentication is essential. With handwriting verification, users are allowed access to the network.
- **Property Security:** Handwriting recognition technology is utilized to authenticate a user's identity and manage the security of the premises.
- **Archival Security:** The material provided by library & archival security mostly relates to security in libraries, archives, and other information centers. This includes data and communications security, physical security, and the examination of related social, legal, and ethical issues. By using one-to-many relationships, the writer identification algorithm finds the most likely match among the writers in the list.

3. WRITER IDENTIFICATION SYSTEM

Writer identification involves determining the authentic author among a group of registered candidates by assessing the resemblance in their handwriting illustrated in Figure 1.

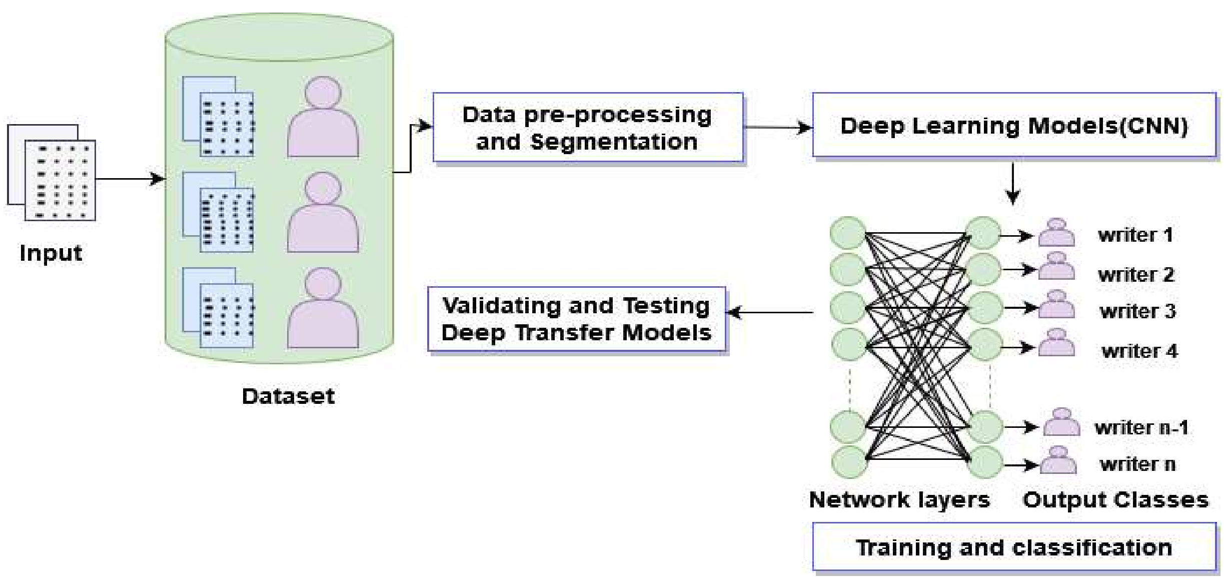

Fig. 1: Block diagram of Writer Identification

ISSN: 2752-3829 Vol. 3 No.2, (December, 2023)

3.1 Classification based on writer Identification System

For offline Writer Identification System, the user writes on a page or sheet, which is subsequently, digitized using either a digital camera or an optical scanner. The input is received by the system as a scanned document and recognizes the text via feature analysis. On-line writer identification: In this approach, the user composes text on a digital tablet or personal digital assistant (PDA), where the system operates in real-time. The dynamic identification system deals with information related to spatial co-ordinates, pressure, and inclination, etc. Writer identification system further categorized as: Text dependent writer identification: The identification approaches extract the feature from the text Image. Text Independent writer identification: The identification process extracts the feature from character shape**.**

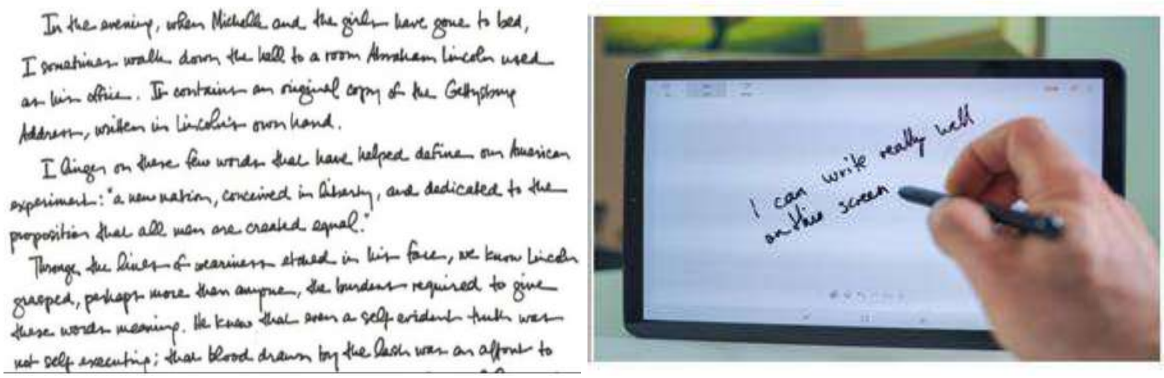

Fig.2: (a) Off-line handwriting (b) On-line handwriting

4. PHASES OF WRITER IDENTIFICATION SYSTEM

The primary aim of this study is to create a writer identification system capable of recognizing online and offline handwritten scripts. The research work includes the following objectives.

1. The initial stage involves gathering data from a range of writers.
2. The digitized data is further used for pre-processing. The pre-processing includes binarization, noise removal, and Histogram modification etc.
3. The image segmentation step includes partitioning of digital image into multiple segments for analysis. Image segmentation serves the purpose of identifying both the objects and their boundaries within an image. This step includes techniques, for instance, line segmentation, word segmentation and character segmentation. Feature extraction step includes extraction of salient features, capture curvature and roundness, texture based and spectral features of handwriting. The off-line feature extraction includes various types, for instance, document feature, paragraph feature, line feature, word feature and character feature.
4. Classification of writer is the last step of the handwriting identification process. The features that have been extracted will be input into a classifier for the purpose of assigning a writer's identification.

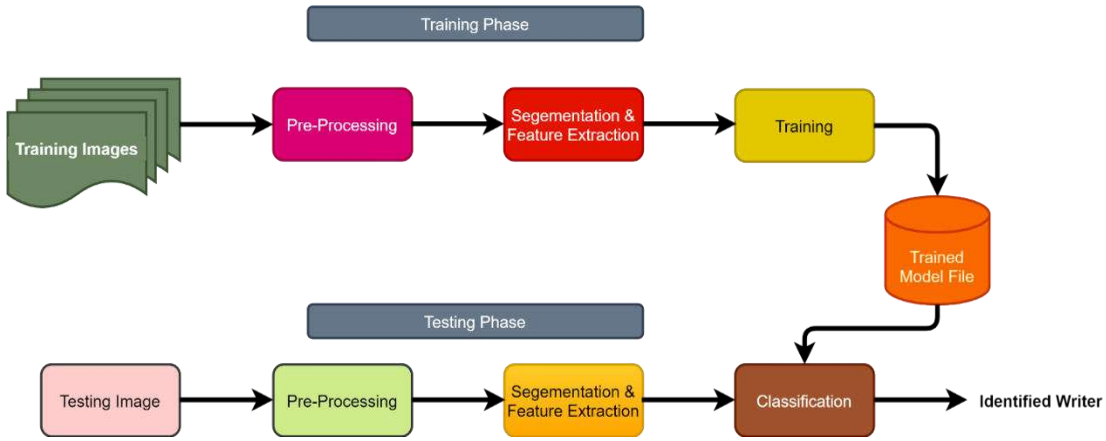

Fig.3: Phases of Writer Identification System

5. REPORTED WORK

Indic script, Non-Indic script, and Deep Learning Based methods are the three subsections that cover the research that has been done to identify the writer of a handwritten text. The study's aim or purpose has been briefly explained before they begin any comparison analysis. Following the general analysis, the methods, results, conclusions, and datasets are provided in tabular form, i.e., from Tables 1, 2, 3,4, and 5.

5.1 Indic Script

India is a nation with a rich tapestry of cultures and religions, characterized by diverse linguistic families. Among these families, languages are primarily categorized into two major groups: Indo-Aryan and Dravidian.

Gurmukhi

Gurmukhi, a script attributed to Guru Angad (1504–1552), the second Sikh Guru, has been adapted, standardized, and is currently the official script for the Punjabi language in the state of Punjab. It is worth noting that Punjabi is also transcribed in the Perso-Arabic Shahmukhi script. Gurmukhi comprises 35 letters, with the initial three being vowels. Kalra et al. [1] implemented a method for identifying writers from offline isolated handwritten Gurmukhi characters. In this approach, collected data of 70% was allocated for training purpose, while as 30% was reserved for testing. Zonal and open endpoint intersection feature extracted from the character images. KNN classifier was employed, the system attained an accuracy of 53%. Moreover, by utilizing the classifier multi-layer perceptron reported an accuracy of 55%. Kumar et al. [2] developed handwriting grading system for writer in an offline mode, which was implemented on data acquired from a pool of 100 writers. The study resolved on offline Gurmukhi character evolution, a range of features such as directionality, zoning, diagonal strokes, intersections, open endpoints, and Zernike moments were incorporated The Bayesian classifiers and Hidden Markov Models (HMM classification techniques implemented in this research. Tests using five widely recognized printed Gurmukhi fonts were used and the system reported 72% accuracy. Verma et al. [3] reported a zone-based character recognition technique for handwritten Gurmukhi script. Data gathered from 10 different writers, generating a total of 4280 image dataset. The study focused to pinpoint the writing zones of distinct characters, classification phase utilized Hidden Markov Model. The experiment attainted an accuracy of 88.4% for both character recognition and writing zone identification based on the Gurmukhi script. Sakshi et al. [4] mentioned a system to identify writer for the Gurmukhi script, a database comprising 49,000 samples from 70 distinct writers were used for experiment. Various features were integrated into the system, such as fitting, zoning, diagonal, power curve, transition, open endpoints, intersection, parabola curve fitting, horizontal, and centroid. To classify the writer, various machine learning algorithms such as Random Forest, Naive Bayes, Decision Tree, and AdaBoostM1 implemented. The reported results for the Gurmukhi script indicate an impressive accuracy rate of 81.75%. Singh et al. [5] mentioned a method to recognize handwritten Gurmukhi character combinations, including matras. To train the classifier, a dataset comprising 52,500-word samples from 175 different writers was utilized. The extracted features consisted of pre-processed x-y points, discrete Fourier transform characteristics, and directional attributes. SVM employed for the classification purpose. The reported accuracy rate of system was 99.75% implemented on dataset which contained 21,500 images of character combinations with matras. Kumar et al. [6] Introduced an approach which used dataset of Gurmukhi characters based on pre-segmentation. This approach relied on employing particular attributes, including zoning, diagonal patterns, transitions, and peak extent. To evaluate its effectiveness, the conducted experiments employing two classifiers: KNN and SVM. The dataset encompassed 31,500 samples from 90 distinct writers. From which 30% of the dataset was applied for testing and rest of the data was utilized for training.

Devanagari

Devanagari script belongs to the broader Brahmic script family and is used across regions such as Nepal, Tibet, and India. The Devanagari script incorporate 47 primary characters, comprising 14 vowels and 33 consonants. Sagar et. al. [7] demonstrated a technique by utilizing the Radon transform and slant analysis. First, a variety of writers' handwritten documents were collected, and then digitized the documents. Dataset CPAR-2012 was utilized, which consists of a diverse collection of Devanagari script samples, including handwritten digits in constrained, semi-constrained, and unconstrained image samples. The dataset also includes isolated fonts, constrained, and unconstrained pangrams. Slant feature was extracted for writer identification. KNN and neural network classifiers were employed for classification task. The study reported promising results, achieving an accuracy rate of 70% while using the KNN classifier. (Kumar et al.) [8] mentioned a dataset for Devanagari document recognition, the study referred benchmark CPAR-2012 dataset. This database encompasses 35,000 isolated handwritten numerals, 83,300 characters, and 2000 constrained and 2000 unconstrained handwritten pangrams. The handwritten images underwent preprocessing to convert them into binary images, effectively eliminating noise to enhance data quality. The KNN classifier attained a commendable identification rate of 82.49% on both the training and test sets. Halder et al. [9] implemented a research on Devanagari characters for the task of writer identification. The dataset contains contributions from 50 unique individuals, totaling 10,750 alphabets in Devanagari, 2,500 numbers, and 3,000 vowels. The feature extraction techniques applied includes contour point analysis, direction coding, and down-sampling. The classifiers implemented for this study includes LIBSVM and LIBLINEAR. Remarkably, System reported an impressive accuracy rate of 99.12%.

Gujarati

Gujarati is a member of the Devanagari language family and notably lacks shirolekha in its numerals. It consists a total of 31 consonants. The Gujarati script is the writing system used for the Gujarati language, which is primarily spoken in the western Indian state of Gujarat and by the Gujarati diaspora worldwide. The Gujarati script is a unique and important writing system used for the Gujarati language, and it plays a crucial role in preserving and promoting the cultural and linguistic heritage of the Gujarati-speaking community. Chokssi et al. [10] introduced an approach to identifying similar appearances among Gujarati characters, implemented the k-NN, Fuzzy-KNN Algorithm, and General Regression Neural Network. The study utilized a dataset consist of 30 samples of every character in bold, italic, and regular. For the evaluation, a combination of structural, geometric, and wavelet features ware extracted. The system’s performance with the General Regression Neural Network achieved 97% accuracy, Fuzzy KNN achieved a perfect 100%, and the K-Nearest Neighbor achieved 67% accuracy. Desai [11] developed a multi-layered feed-forward NN for classification. Pre-processing step involved thinning and skew correction procedures. The system reported an accuracy of 86.66% in identifying Gujarati handwritten digits. The training and testing databases contained 610 and 2,650 digits, respectively.

Malayalam

The Malayalam script originates from the South Indian state. Malayalam script is the writing system used for the Malayalam language, primarily spoken in Kerala (India) and the Union Territory of Lakshadweep, as well as by Malayali communities around the world. It is a syllabic alphabet that belongs to the Brahmic family of scripts, which also includes scripts like Devanagari, Tamil, and Kannada. Within the Malayalam character set, there are a total of 73 characters, comprising 13 vowels, 5 chillu characters, 36 consonants, 4 consonant signs, and 12 vowel signs. Rahiman et al. [12] developed a backpropagation neural network technique for printed character recognition. The approach Wavelet multi-resolution analysis make use of wavelet transform features. A Feed Forward Backpropagation Neural Network was used to accomplish the character recognition tests. There were 715 images in the collection in all, 325 of which had noise and 400 of which did not. The system attained remarkable 92.0% accuracy rate. Idicula et al. [13] created a method for utilizing graphemes to identify writers in Malayalam. Graphemes are microscopic writing units that are taken from handwritten manuscripts. Each writer's graphemes have unique characteristics and significant patterns. The dataset included contributions from two hundred and eight writers, whose page counts varied. For example, 50 writers submitted a single page, 215 writers wrote two pages or more, and 15 writers submitted four pages or more. The system showcased a remarkable accuracy rate of 89.28% Manjusha et al. [14] mentioned handwritten character image database for the Malayalam language script, comprising 85 distinct characters contributed by 77 different writers. The selection of classes related to these Malayalam character was based on the script's unique orthographic characteristics. To extract isolated handwritten characters from this dataset a segmentation algorithm named as an active contour model-based image applied. The resultant character images were then subjected to various feature extraction techniques, including vectorized representations of pixel intensity values, localized features, and curvature features. The techniques used for feature extraction included run length counting, curvelet transform, singular value decomposition, histogram of oriented gradients, picture pixel analysis, and Gabor filters. Through extensive experimentation, the scattering convolutional network-based feature descriptors demonstrated exceptional performance, achieving a remarkable recognition accuracy of 91.05%.

Oriya

An early version of the Bengali script served as the model for the Oriya script, belonging to the Northern group of South Asian scripts. It is employed to transcribe the Oriya language, primarily spoken in the contemporary Indian state of Odisha, situated on the eastern coast of India. The Oriya script consists of 64 characters, encompassing both consonants and vowels. Chanda et al. [15] designed an identification system which was text independent for the Oriya script. The research focused on leveraging directional chain codes and curvature-based features. To classify the writers, Support Vector Machines (SVM) model was employed. The system yielded a notable accuracy of 78.50%. For the practical implementation of the system, a comprehensive dataset featuring contributions from a diverse group of 100 writers was collected. Misra et al. [16] Devised a method for enhancing Oriya handwritten character recognition by implementing contour-based features. By Employing the combination of Backpropagation neural network and Hidden Markov Model as the classification tools, the system achieved an impressive accuracy rate of 84.5%. The dataset comprised a substantial 17,000 samples contributed by 150 different writers. Panda et al. [17] introduced an algorithm for the offline typewritten Odia characters recognition. The database encompasses a comprehensive collection of Odia script's character images. They utilized a method that combined Unicode Mapping with Template Matching. The system achieved a 76.45% accuracy rate by implementing this template matching technique in conjunction with classification, Unicode mapping, along with the direct vector matrix.

Tamil

Tamil, a Dravidian language, is primarily spoken by the Tamil community in India and also by the Tamil diaspora, including Sri Lankan Moors, Douglas, and Chindians. It holds the status of an official language in two countries: Singapore and Sri Lanka. Furthermore, Tamil is the official language of the Indian state of Tamil Nadu. Remarkably, Tamil stands as one of the world's enduring classical languages. It encompasses 12 vowels, 18 consonants, and an impressive 216 compound characters formed by combining vowels and consonants. Jayanthi et al. [18] Devised a method to identify the author of Tamil handwritten data on the basis texture analysis. This method relies on extracting distinctive characteristics from the GLCM derived from scanned images. The dataset encompasses handwritten samples from 70 different writers, achieved an accuracy rate of 82.6%. Thendral et al. [19] Introduced a text-dependent method for identifying writers based on the Tamil script, leveraging both global and local features. The classification technique employed was the Decision Tree method, accompanied by a 10-

fold cross-validation procedure. Impressively, system reported remarkable accuracy of 98.6%. Thendral et al. [20] stated a discriminative approach, with pooled features of handwriting. The classification method employed was the SVM. The dataset comprises of 100 Tamil images, each of which has been transcribed by 300 distinct writers. Thendral et al. [21] Using offline handwritten text, a writer identification system was developed for the Tamil script. The 500 words in the dataset were written by ten different authors. For the purpose of classification, SVM was utilized in conjunction with the design of the Basis Function (RBF) model. The method showed a 90.6% identification accuracy.

Telugu

The Telugu language, which is predominantly spoken in the Indian states of Andhra Pradesh and Telangana, is written in the Telugu script, a Brahmic script. Each script unit in the Telugu writing system is made up of a vowel and a consonant, making it an Abugida or segmental writing system. There are 56 fundamental characters in the Telugu script, comprising both vowels and consonants. Purkait et al. [22] presented a method to use directional morphological traits to improve the writer identification system for handwritten Telugu images. K-curvature, directional erosion, directional opening, and directional curvature features were retrieved. The Nearest Neighbor method was used for categorization. The main reason this system was able to attain an accuracy rate of 71.73% was because the directional opening strategy worked so well. Andrew et al. [23] developed a technique to recognize writers using Telugu handwritten documents. In order to make this easier, a brand-new dataset called the IIITSTHDB was created. It consists of 150 writers' handwritten Telugu documents. Directional filters were used to extract features based on descriptive convolution. Useful classifiers for classification phase included such as SVM with an RBF kernel, SVM with a linear kernel, and Nearest Neighbor. The reported accuracy rate of the system was 71.0%.

Kannada

Karnataka is a state in southern India where Kannada is the official language. The language based on Dravidian languages, which is prominent in various parts of India. The Kannada script character set, used for writing the Kannada language, is a syllabic alphabet with 49 basic characters. Mukarambie et al. [24] mentioned technique for identifying writers through texture analysis of Kannada handwriting in this study. The researcher assembled a dataset comprising handwriting samples from 20 writers. The image preprocessing steps involved eliminating non-text regions, correcting skew, eliminating noise, and binarizing the images. Various distinctive features were extracted from the diverse handwriting samples, primarily using the discrete cosine transform (DCT), Gabor filtering, and gray level co-occurrence matrix (GLCM). For classification, a K-nearest neighbors (KNN) was employed. It was observed that the features derived from Gabor filtering were more accurate in writer identification compared to GLCMs and DCT, and the system achieved an impressive accuracy rate of 88.5%. The Comparative Analysis of Indic Scripts are mentioned in Table 1.

Table 1: Writer Identification Work in Indic Scripts.

Author	Script	Dataset	Feature extraction method	Classifier	Accuracy
kalra et al. [1]	Gurmukhi	30 Writers	Zoning features, open end point, and intersection	Multilayer Perceptron Network and KNN	53%
Kumar et al. [6]		90 writers	Zoning, diagonal, transitional and peak extend	SVM	89.85%
Kumar et al.[2]		100 writers	Directional, zoning, diagonal, intersection and Zernike moments	Hidden markov model, Bayesian classifier	93.50% 89.93%
Verma et al.		10 writers	Upper, middle and lower zones	Hidden Markov	88.4%

[3]				Model	
Singh et al. [5]		175 writers	discrete Fourier transform, Pre-processed x-y points, as well as directional features	SVM	99.75%
Sagar et al. [7]	Devnagari	CPAR-2012	Radon transform, Hough transform, Zernike moments	KNN,NN	70%
Kumar[8]		CPAR-2012	Ostu method, Noise removal and Morphological operations	KNN	82.49%
Halder et al. [9]		50 writers	Point, direction code, down sampling	LIBSVM and LIBLINEAR	99.12%
Choksi et al. [10]	Guajarati	30 samples taken from each character	Structural and wavelet features	Fuzzy KNN	100%
Desai [11]		199 writers	Aspect ratio, extent of alphabet and image sub-division	KNN	86.66%
Rahiman et al.[12]	Malayalam	715 Images	Wavelet transform	Back propagation Neural network	92%
Idicula el al. [13]		280 Writers	Graphemes	Nearest neighbor	89.28%
Manjusha et al.[14]		77 writers	Curvelet transform, Gabor filters, run length count.	Scattering Convolutional network	91.05%
Chanda et al.[15]	Oriya	100 writers	Curvature based and Directional chain code	SVM	94.00%
Mishra et al.[16]		150 writers	Contour based	Hidden Markov Model	84.5%
Panda et al.[17]			Vector Matrix	Template Matching	97%
Jayanthi et al.[18]	Tamil	70 writers	Gray level co-occurrence matrix	Feature Vector	82.8%
Thendral et al.[19]		10 writers	Local and Global features	Decision tree	98.6%
Thendral et al. [20]		300 writers	Pooled features	Support Vector Machine	100%
Thendral et al.[21]		10 writers	Local and Global features	Support Vector Machine and Radial Function	90.6%
Purkait et al.[22]	Telugu	22 writers	Directional Opening, k-curvature, Directional erosion	Nearest Neighbor	71.73%
Andrew et al. [23]		150 writers	Directional filters for Descriptive convolution based features	Support Vector Machine and K Nearest Neighbor	71%
Mukarambi	Kannada	20 writers	Gabor filtering and gray level	KNN	88.5%

et al.[24]			features		

In addition to this, the graphical analysis of the afore mentioned table have been also presented in figure 4,5,6, and 7.

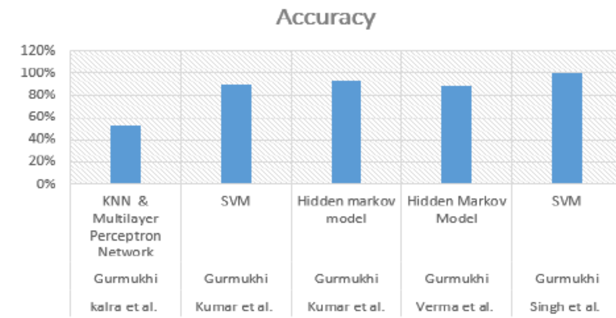

Fig 4:Graphically Analysis of Gurmukhi Script

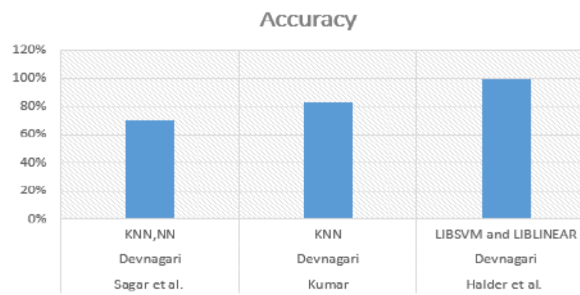

Fig 5:Graphically Analysis of Devanagari Script

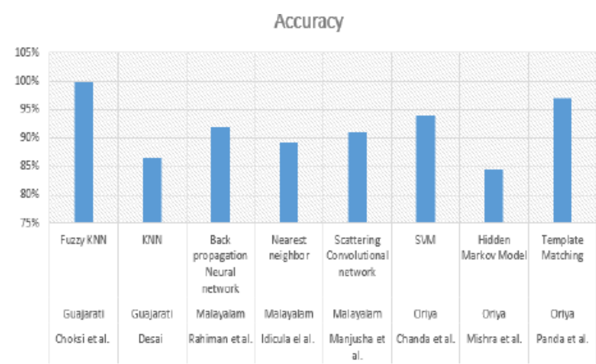

Fig 6: Graphically Analysis of Guajarati, Malayalam, Oriya Script

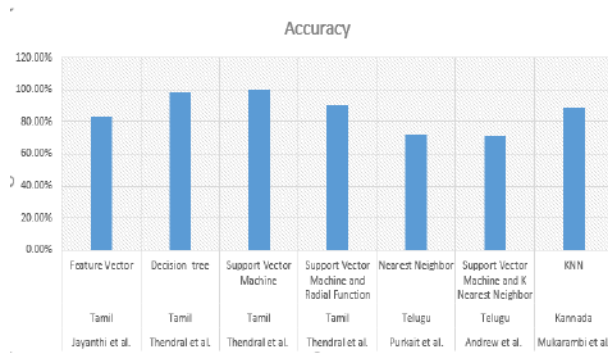

Fig 7: Graphically Analysis of Tamil, Telugu, Kannada Script

5.2 Non-Indic Script

Various non-Indic scripts are utilized across the globe, each exhibiting distinct characteristics depending on the country of use. Among these non-Indic scripts are English, Arabic, French, and many more.

Arabic

Arabic stands as one of the world's prominent languages, holding a place within the top six. As the sacred language of Islam's holy book, the Qur'an, it enjoys extensive usage across the Muslim globe. This language is part of the Semitic language family, that also encompasses Amharic and Hebrew, the primary language of Ethiopia. Arabic's script is uniquely characterized by its "right-to-left" direction and cursive style, composed of a set of 28 distinct letters. Asi et al. [25] implemented an algorithm for identifying writers in Arabic historical manuscripts. The experiments were conducted on two benchmark datasets: WAHD and KHATT. The WAHD dataset consists of 353 manuscripts sourced from two primary locations. It includes 333 manuscripts from the Islamic heritage project and an additional 20 manuscripts from the National Library in Jerusalem. On the other hand, KHATT is a contemporary Arabic handwritten text dataset, compiled from the work of 1000 scribes originating from various countries. The development of KHATT was a collaborative effort involving research groups from Arabia, TU Dortmund in Germany, and Saudi Arabia. For feature extraction method the Curvature and Roundness were employed. The classification phase of algorithm involved two methods: Averaging and weighted voting. The algorithm achieved noteworthy results in the experiments. An accuracy of 83% obtained using the WAHD dataset and an impressive 88.9% accuracy rate with the KHATT dataset. Chaabouni et al. [26] mentioned an experiment which was performed on samples of 110 writers from the ADAB (Arabic) database, aimed to examine the potential of extracting multi-fractal attributes for the purpose of differentiating individuals based on both online and offline writing. Remarkably, the results showed a substantial identification rate of 83.8%. Future research will expanded to encompass larger databases and diverse scripts, such as English, Chinese, and more.

French

French serves as the official language not only within France but also in certain regions of Belgium and Switzerland, as well as in Monaco and various areas of Canada. Moreover, it holds the status of an official language or a secondary language in 55 countries globally and is renowned as the second most prevalent foreign language employed for international communication, following English. Nearly 300 million individuals speak French either as their mother tongue or as a second language. The French language comprises 28 letters in its alphabet. Tan et al. [27] implemented an automated text-independent framework for writer identification, with a primary research emphasis on segmentation at the character level based on online handwritten documents. The dataset comprises samples from 120 writers using the French script. Employing the Fuzzy C-means method, the system computed an impressive accuracy rate of 99.2%. Importantly, the proposed technique demonstrates its efficacy in handling documents written in English, Italian, or German, all of which belong to the Latin script family, just like French.

Chinese

Chinese is a part of the Sino-Tibetan language family, that is an important linguistic group that is related to the Indo-European family, which also includes languages like Hindi, German, and French. With over a billion speakers spread over Southeast Asia, South Asia, and East Asia, the Sino-Tibetan language group stretches from northeastern India to northeastern China. Chinese is not a single language; rather, it belongs to a family of languages, much to the Romance languages, which also include French, Spanish, Italian, Romanian, and Swiss Romansch. Chinese characters are symbolic in nature. There is a large repertoire of over 50,000 characters that may be conveyed in block script or cursive handwriting. Shin et al. [28] mentioned two algorithms for writer identification task. The first algorithm, the Block Type Model focused on analyzing the positional relationships and shapes of Chinese characters. In contrast, the second algorithm, Hidden Feature Analysis utilized a multi-parameter approach that considers all the strokes in identifying the writer. Remarkably, these methods have demonstrated remarkable accuracy, with just eight Chinese characters being sufficient to obtain 99% accuracy rate. Li et al. [29] introduced a feature based on histogram for identifying Chinese writers. The HIT-MW Chinese handwriting database was compiled, featuring contributions from 240 writers. Pre-processing utilized the Otsu technique for binarization and the application of Sobel operators for edge detection. Microstructure features were subsequently extracted and compared with two established histogram-based features: the contour hinge features and the multi-scale contour-hinge. The system obtained an identification accuracy of 99.6%.

English

English stands as the most widely spoken language, with a vast number of speakers, and it ranks as the third most commonly spoken native language across the globe. The English language comprises a total of 26 characters. Bertolini et al. [30] used textural descriptors for the task of writer identification and verification. A classification approach based on dissimilarity representation was used to handle verification issues. Apart from the assessment of two different texture descriptors, which were local binary patterns and local phase quantization, the study also explored important issues related to dissimilarity representation. These included investigating the influence of the number of references used for verification and identification, assessing the system's performance in writer identification, and comparing the dissimilarity-based approach with other feature-based strategies. Two diverse datasets were leveraged for the experimental phase: the IAM database and The Brazilian forensic letters dataset. The proposed approach utilized LPQ features which demonstrated accuracy rates of 96.7% and 99.2% for the BFL and IAM databases, respectively. Fiel et al. [31] introduced a method based on local features, notable for its independence from a binarization step. Firstly, this approach involved the computation of local features of the image. These features were then employed in conjunction with a predefined codebook to generate an occurrence histogram. Then histogram utilized to find out the identity of the writer or measure the similarity to other handwritten documents. The performance of this approach was assessed using two different databases: first one was the IAM encompassing 650 writers, and the second being the TrigraphSlant featuring 47 writers. For classification phase, a Nearest Neighbor classifier utilized, which obtained a remarkable accuracy rate of 97.5% for the IAM dataset and an even higher accuracy of 98.9% for the TrigraphSlant dataset Siddiqi et al. [32] developed an efficient technique conducted using the IAM dataset. A two-stage technique was devised for the identification procedure. The initial one involved categorizing the handwritten text into various classes based on writing styles, while the later stage aimed to identify the author of an unknown document. In the first phase, Gabor filters were utilized to analyze the handwriting and segregate it into different classes. In the second phase, the handwritten content was further dissected into numerous small sub-images, after that sub-images with morphological similarities were grouped together. For precise identification a Bayesian classifier was applied to the subset of this data. The system's performance was evaluated using a set of 50 documents written by the same individuals. Impressively, the system achieved 94% identification rate, showcasing its effectiveness in writer identification task. Zhu et al. [41] mentioned a writer identification system for text-independent offline that makes use of a sparse auto-encoder codebook. An unsupervised technique, the sparse auto-encoder for feature learning, played a pivotal role in this method. The scheme was thoughtfully designed to generate SAEC codebook be language-agnostic, enabling its adaptability of languages to a wide range, even when dealing with multiple languages. A feature of this algorithm was its ability to perform without the requirement of text segmentation. To validate its accuracy, a series of experiments were conducted on two different datasets, one in Chinese (HITMW) and the other in English (IAM). The identification performance was remarkably impressive, provided an accuracy of 98.59% for Chinese and 99.17% for English. Furthermore, the research delved into an evaluation of various parameters that reflect the performance of the identification system. These feature included the factors like patch size, text quantity, and the number of patches utilized in the sparse auto-encoder. For the classification phase, the K-nearest neighbors (KNN) algorithm was utilized. Pandey et al. [43] introduced handwriting recognition system that was based on grapheme. The system involved segmentation of handwritten images into individual graphemes. Handwritten images were segmented into separate graphemes by the algorithm. These graphemes were encoded by using horizontal projection profile that logs the location of the first foreground pixel in each row of the bounding box. The K-means clustering approach was employed to generate a reference set. The model's performance was evaluated, obtained an accuracy rate of 88.57%. Bertolini et al. [44] reported an experiment which was conducted using the QUWI dataset. Two distinct texture descriptors, namely Local Binary Patterns (LBP) and Local Phase Quantization (LPQ) were utilized for extracting the feature. The SVM classifier was then employed for the purpose of classification, obtained an accuracy rate of 70%. Venugopal et al. [47] developed a system for online writer identification, which did not rely on specific text information. The experiments were performed on the IAM and IBM UB 1 datasets and the goal of this study was to enhance the existing Vector of Local Aggregate Descriptor (VLAD). In this approach point-based features were extracted for writer identification. For the classification phase, Support Vector Machines was employed. Remarkably, the system reported an accuracy of 97.81% on the IAM dataset and 94.37% on the IBM UB 1 dataset, demonstrating the effectiveness of the descriptor and methodology in the domain of online writer identification. U et al. [48] mentioned HMM-based system outperformed than GMM-based system. Experiments performed at the line and paragraph levels using the IAM Online English Handwritten Text Database (IAM-OnDB). This dataset consists of online data acquired from a whiteboard. The study focused on enhancing the GMM-based system's performance by integrating temporal information. A sequence of feature vectors was extracted and utilized to train the classifier, which was based on the Hidden Markov Model (HMM). This sophisticated approach provided an accuracy rate of 94.5%. Garz et al. [49] introduced a set of writer identification descriptors, offering a straightforward and efficient computation procedure. The experiments performed on four datasets: IAM, ICDAR’11C, ICDAR’11F, and ICDAR’13. These descriptors such as strokes, junctions, endings, and loops derived. The reported identification rates were 86.1%, 98.6%, 86.7%, and 84.1%. Wu et al. [50] developed a methodology for automated writer identification system based on offline text-independent using Scale Invariant

Feature Transform (SIFT). The approach leverages two different SIFT features such as the SIFT Descriptor Signature (SDS) and the Scale and Orientation Histogram (SOH. The experiments were conducted on six diverse publicly available datasets: IAM, Firemaker, U-HIT-MW, L-HIT-MW, ICDAR2011, and ICFHR2012. The system achieved impressive accuracy rates of 98.5%, 92.4%, 95.4%, 99.5%, and 98%, respectively, on the mentioned databases. Hu et al. [51] focused on offline text-independent identification of Chinese handwriting writers. Experiments were conducted on a recently curated dataset, known as CASIA Offline DB 2.1, which consists of handwritten samples from 240 distinct writers. The approach employed the scale-invariant transform features (SIFT) descriptors. These descriptors extracted localized directional information from the Chinese characters in the database. K-nearest-neighbor classifier was employed and obtained a notable accuracy rate of 96.25%. Fecker et al. [52] mentioned variety of techniques for extracting features and classifying historical Arabic manuscript. These approaches were tested using a publicly accessible database comprising 60 distinct manuscripts originating from the 13th century to the early 20th century. These manuscripts were conveniently accessible online as part of the Islamic Heritage Project (IHP). The feature extraction method included the contour, textural, and key point-based, while the classification strategies involved averaging and voting. The obtained accuracy of system was 92.5%, which was attained by leveraging a key point descriptor-based feature extraction technique. Christlein et al. [53] Utilized GMM super vectors for encoding the feature distribution of individual writers. The experiments were conducted on the ICDAR2013, CVL, and KHATT databases. The parameters derived from the adapted GMM models were aggregated to model the GMM super vector. The resulting accuracy rates were 89.4% for ICDAR2013, 91.0% for CVL, and 97.2% for KHATT. Newell et al. [54] mentioned two approaches in the study: the oriented Basic Image Feature Columns (oBIF Column) encoding scheme and the Delta encoding method. These methods were performed on the IAM dataset and ICDAR 2012. The Nearest Neighbor algorithm with the Euclidean distance as the classifier was utilized for evaluation and achieved an impressive accuracy of 99% when implemented on 300 writers of IAM dataset and a substantial 93.1% accuracy for the ICDAR 2012 dataset. He et al. [55] stated two distinct strategies for offline text-dependent and text-independent. The research involved the evaluation of 100 Chinese handwritten documents contributed by 50 different writers. Feature extraction was accomplished using Gabor filters and autocorrelations. The extracted features were executed using a Weighted Euclidean Distance classifier. The outcomes reported a 96% accuracy rate for the text-dependent method and a 90% accuracy rate for the text-independent approach. Khalifa et al. [56] introduced an approach to enhance the codebook model and conducted experiments on both the ICFHR 2012 and IAM datasets. Structural features were extracted, and a classifier known as Kernel Discriminant Analysis with Spectral Regression (SR-KDA) was implemented. The achieved accuracy rates for IAM was 92% and ICFHR 2012 96%. Christlein et al. [57] stated a method to enhance offline writer identification by ensuring robustness. This method was evaluated on three well-known public datasets, namely ICDAR, CVL, and KHATT. To capture the unique handwriting feature, the technique employed GMM super vectors as an encoding mechanism. The obtained accuracy levels of 99.3%, 98.0%, and 97.2%. Hannad et al. [58] mentioned the handwritten samples were broken into smaller chunks, each of which was given its own unique texture. These segments were examined in order to derive texture descriptors, including local phase quantization (LPQ), local binary patterns (LBP), and local ternary patterns (LTP) histograms. This technique was performed on datasets from the IFN/ENIT and IAM that included handwritten text in both Arabic and English. A Hamming distance model was implemented as the classifier. The reported results revealed a high level of accuracy rate of 94.89% for IFN/ENIT and 89.54% for IAM. Chahi et al. [59] introduced Block-Wise Local Binary Count (BW-LBC) for Offline Text-Independent Writer Identification. The experiment performed on datasets such as IAM, IFN/ENIT, AHTID/MW, and CVL. K-Nearest Neighbors (KNN) classifier with Hamming distance as the similarity metric was employed. The reported accuracy rates for these diverse datasets were as follows: 88.99%, 96.47%, 99.53%, and 98.38%. Chahi et al. [61] developed LSTP to acquire the complex local writing structures existing in certain writing regions, which were called "connected components." In the classification phase, nearest neighbour (NN) based on Hamming distance was used to compare and match LSTP feature vectors. This system used seven well-known handwriting standards (English CVL, Arabic IFN/ENIT, Dutch Firemaker, English-Chinese CERUG, English-Greek ICDAR2013, and Hybrid-language ICDAR2011). The reported efficacy was as: IFN/ENIT: 98.28%, CVL: 99.83%, English and Greek ICDAR2013: 98%, IAM: 96.80%, Firemaker: 98%, Hybrid ICDAR2011: 100%, CERUG-CN: 100%, CERUG-EN: 98.09%. Pandey et al. [63] mentioned the Allographic features at the character or grapheme level from a segment. Features were extracted from windows or segments defind from the word image. These features were subsequently organized into clusters using K-means. When applied to handwritten words written by ten English authors in a dataset, the system achieved a notable writer identification accuracy of approximately 63%. The comparative analysis of Non-Indic Scripts are mentioned in table 2.

Table 2: Writer Identification Work in Non-Indic Scripts.

Author	Script	Dataset	Feature Extraction Method	Classifier	Accuracy
Asi et al.[25]	Arabic	KHATT& WAHD	Curvature and Roundness	Averagin g and Weighted voting	83% 88.9%
Chaabouni et al. [26]	Arabic	110writers	Multi-fractal features	K Nearest Neighbor	83.8%
Tan et al.[27]	French	120 Writers	Character Level features	Fuzzy C-mean kernel	99.2%
Shin et al.[28]	Chinese	48writers	Inter and Intra stroke	Block Type Model	99%
Li et al.[29]	Chinese	240writers	Histogram based feature	Weighted Chi-Square Metric	99.6%
Bertolini et al.[30]	BFL &IAM	315 Writers 650 Writers	LBQ and LPQ	Support Vector Machine	96.7% 99.2%
Siddiqi et al.[31]	IAM and TrigraphSlant	650 writers 47 writers	Slant and codebook	Nearest Neighbor	97.5% 98.9%
Siddiqi et al.[32]	IAM	650 writers	Inherent features	Bayesian Classifier	94%
Zhu et al.[41]	English and Chinese	IAM and HIT-MW	Sparse auto-encoder (SAE) based codebook (SAEC) and Non-text-segmentation feature	KNN	98.59% 99.17%
Pandey et al.[43]	English	IAM	Graphemes	KNN	88.57%
Bertolini et al.[44]	Arabic and English	QUWI	Local Binary Patterns as well as Local Phase Quantization Discriminative features	SVM	70%
Durou et al.[46]	English Arabic	IAM ICFHR- 2012	Oriented Basic Image feature with	KNN	92% 97%

			grapheme		
Venugopal et al.[47]	English	IAM online IBM UB1	Point based features	SVM	97.81% 94.37%
Wu et al.[48]	English	IAM-OnDB	Sequence of feature vectors	HMM	94.5%
Garz et al. [49]	English	IAM ICDAR'11C,ICDAR'11F,andICDAR'13	Strokes, Junctions, endings and loops.	Naive Bayes	86.1%, 98.6%, 86.7%, 84.1%.
Wu et al. [50]	English,	IAM, Firemaker, HIT-MW, ICDAR 2011, ICFHR 2012	SIFT Descriptor Signature (SDS) and Scale and Orientation Histogram (SOH)	Manhattan Chi-square	98.5% 92.4% 95.4% 99.5% 98.0%
Hu et al. [51]	Chinese	CASIA Offline DB 2.1	Scale Invariant Transform Feature(SIFT)	KNN	96.25%
Fecker et al. [52]	Arabic, Persian, Turkish	Islamic Heritage project (IHP)	Contour, Textural- and Key Point-based	Chi-square	92.5%
Christlein et al. [53]	German and English	ICDAR-2013 CVL KHATT	GMM super vector	Support Vector Machine	91.0% 89.4% 97.2%
Newell et al. [54]	English	IAM ICDAR2012.a	Oriented Basic Image Feature Columns encoding and the Delta encoding(Δ)	Nearest Neighbor with Euclidean distance	99% 93.1%
He et al. [55]	Chinese	50 writers	Gabor filters and Autocorrelations	Weighted Euclidean Distance	96% 90%
Khalifa et al. [56]	English and Arabic	IAM andICFHR 2012	Structural features	Nearest-neighbor	92% 96%
Christlein et al. [57]	Born digital images real scene images and real scene videos in ICDAR13 and German, English	ICDAR 2013 CVL and KHATT	GMM Super Vectors	Exemplar-SVM	99.3% 98.0% 97.2%

	Text in CVL				
Hannad et al. [58]	Arabic English	IFN/ENIT IAM	Local Ternary Patterns Local Binary Patterns, and Local Phase Quantization	Hamming Distance	94.89% 89.54%
Chahi et al. [59]	Arabic, English, Arabic-German And English	IFN/ENIT IAM AHTID/MW CVL	Block Wise Local Binary Count	KNN with hamming distance	88.99% 96.47% 99.53% 98.38%
Chahi et al. [61]	Dutch, English, Arabic, Hybrid-language, Arabic-English, English-Chinese, English-Greek	Firemaker, CVL, IFN/ENIT, ICDAR2011, IAM, CERUG, ICDAR2013	LSTP feature	Hamming distance NN	98 %, 99.83 %, 98.28 %, 100 %. 96.80 %, 94.28 %. 98%
Pandey et al.[63]	English	10 writers	Character- or grapheme-level allographic features	K-means	63 %

Moreover, figures 8 and 9 provide the graphical examination of the previously given table.

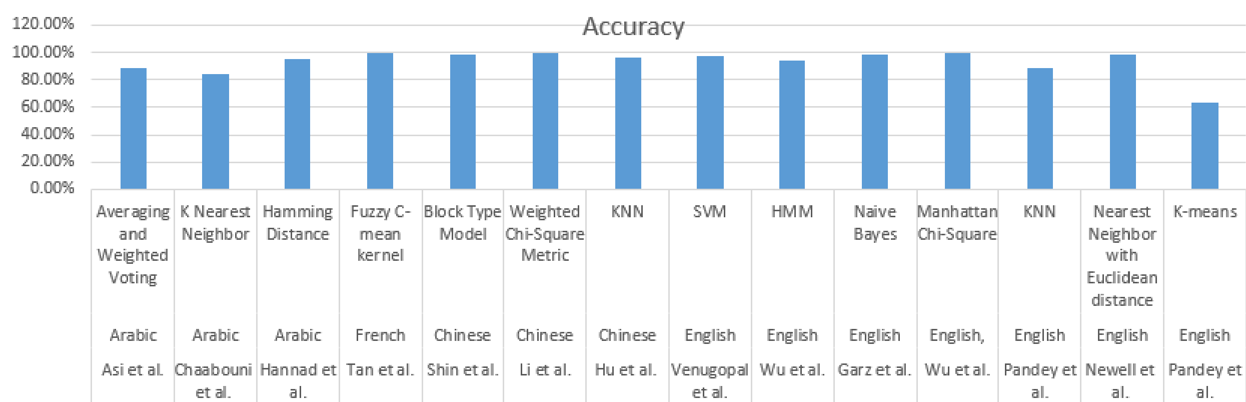

Fig 8: Graphically Analysis of Arabic, French, Chinese, English Script

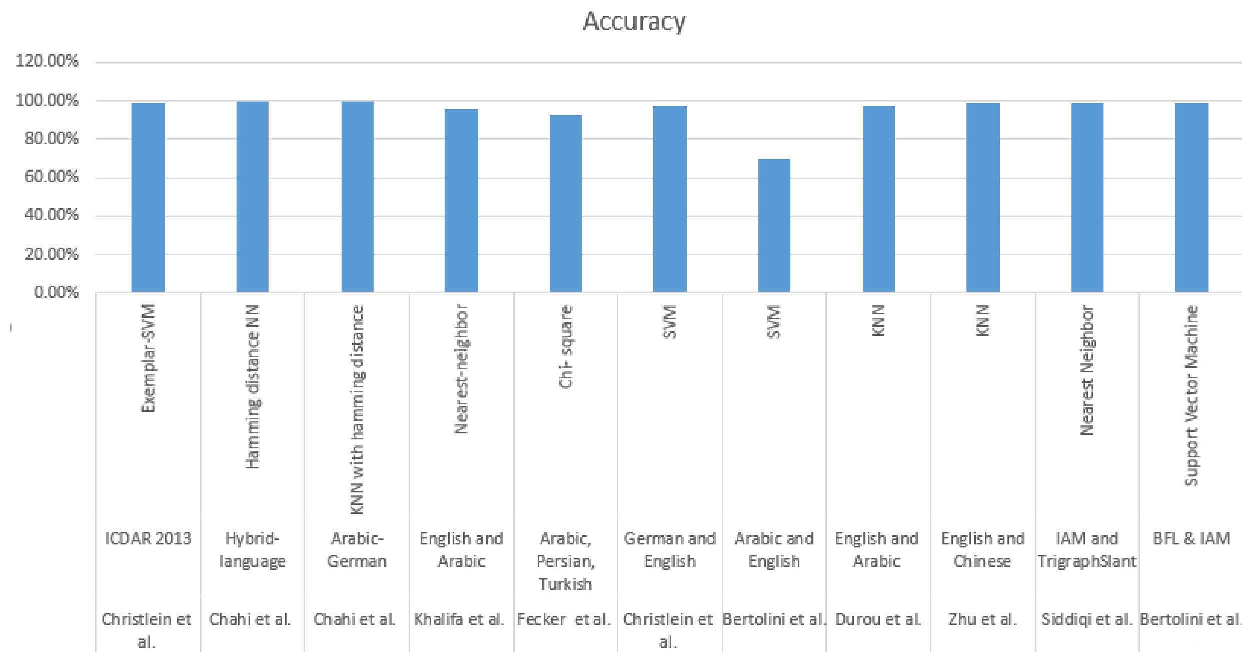

Fig 9: Graphically Analysis of Multilingual Script

5.3 Writer identification based on Deep Learning Approaches

The findings related to writer identification using deep learning methodologies which include a variety of approaches, strategies, and classifiers are shown in this section of the report. Christlein et al. [35] Utilized activation features from Convolutional Neural Networks (CNNs) as localized descriptors for writer identification. This approach was performed on two databases: the ICDAR 2013 benchmark and the CVL. A global descriptor was created by using a Gaussian Mixture Model (GMM) super vector encoding. The outcome of this method was enhanced through normalization with the KL-Kernel. The proposed approach demonstrated impressive results, reported an accuracy of 88.6% on the ICDAR 2013 database and an even more remarkable 97.8% on the CVL database. Chen et al. [36] presented a semi-supervised approach to enhance writer identification through feature learning. A novel method called "weighted label smoothing regularization" was used to augment the data, assigning weighted uniform label distributions to unlabeled data. The research was implemented on the ICDAR2013 and CVL datasets. The baseline convolutional neural network was regularized by using the weighted uniform smoothing regularization approach, which allowed it to learn more unique features that collect the features of different writing styles. Consequently, the system's accuracy on both datasets was outstanding 91.8% and 98%, respectively. Christlein et al. [37] implemented a method for encoding local descriptors forming a global representation suitable for comparative purposes. A comparison of the outcomes of VLAD encoding with triangulation embedding performed and concluded that max-pooling exhibited a slight advantage over sum pooling. While VLAD encoding applied with power normalization, it yielded slightly superior results. The mentioned combination of deep CNN activation features, VLAD encoding, normalization, and Exemplar SVMs displayed consistent performance across three publicly available datasets and indicated an outcome of 93.2%, 98.4%, and 98.0%. Tang et al. [38] implemented a Convolutional Neural Network (CNN) in conjunction with Joint Bayesian analysis. The mentioned approach was put to the test using two widely recognized benchmark datasets, ICDAR2013 and CVL, and organized into two core phases: feature extraction and writer identification. For augmentation of the available data, a customized technique was employed, which enabled the generation of numerous handwriting images for each individual writer. Next phase extracted the CNN-based features, this approach learned features from the entirety of the handwriting images and directly employed them for writer identification. The system performance reached 99.1% accuracy on one benchmark dataset and 93% on the other, solidifying its proficiency in identification task based on text-independent. Rehman et al. [39] Mentioned the

CNN's pre-trained AlexNet architecture on many representations of text-line image patches. The experiment was carried on QUWI dataset. Skew detection, skew correction, normalization, and segmentation, an overlapping sliding window technique were all part of the preparation pipeline. Data augmentation techniques such as contour extraction, negative transformation, and sharpness enhancement applied to generate a substantial amount of data for training deep models. The pre-trained AlexNet architecture was then employed to extract distinctive visual features from these multiple representations of image patches. Then extracted features were input into a Support Vector Machine (SVM) for classification. The attained accuracy rates using the freeze Conv5 layer were as 92.78% for English, 92.20% for Arabic, and 88.11% for a combined dataset containing both Arabic and English text-line images. Yang et al. [40] Developed an approach DeepWriterID, to address the challenge of online text-independent writer identification, focused on the CASIA-OLHWDB1.0 dataset. The path-signature feature was employed and presented technology named DropSegment to facilitate the training of robust convolutional neural networks (CNNs) with limited data. This approach successfully attained an impressive writer identification accuracy of 95.52%. Fiel et al. [42] discovered an approach by utilizing Convolutional Neural Networks (CNN) to generate distinctive feature vectors for each writer. The $\chi 2$-distance metric was used to compare these feature vectors, and preprocessing steps included binarization, normalization, and the rectification of text line skew. This technique's efficacy was evaluated on three distinct databases: 2013 writer identification contests and the ICDAR 2011, along with the CVL dataset. The K-Nearest Neighbors (KNN) classifier was implemented and reported accuracy scores were reaching 98.6%, 98.3%, and 98.3%. Xing et al. [45] Stated data-driven technique which used a Deep Convolutional Neural Network to extract distinctive features from patches of handwritten text. DeepWriter model took localized segments of handwritten text as its input and was trained using softmax classification. To enhance the system performance a multi-stream structure and data augmentation method was employed. Moreover, a patch scanning technique was implemented to effectively handle text images with varying lengths. For the IAM and HWDB datasets, the results provided accuracy rates of 99.01% and 93.85%, respectively. Nasuno et al. [60] Mentioned that Japanese dataset containing 100 unique words contributed by 100 different writers used for experiment. The training dataset consists of 90 of these words while the testing dataset involved the remaining 10 words. Feature extraction was computed by incorpated the Alexnet (CNN), obtained accuracy rate of approximately 90%. Kumar et al. [62] Demonstrated the (SEG-WI) a segmentation-free model, eliminating the need for segmentation and preprocessing steps. Convolutional Neural Networks (CNNs) incorporate to capture feature-level representations of various regions of varying sizes. During testing stage, the model utilized a weighted voting mechanism to amalgamate the knowledge from these selected regions for writer classification, and during training, it computed the overall network loss. Utilizing Deep Convolutional Neural Networks (DCNNs), the model autonomously identified writer-specific features in independent offline text, achieving identification rates of 92.79%, 99.35%, 98.30%, and 100.00%, and 87.06% for the IAM, CVL, IFN/ENIT, Kannada, and Devanagari datasets, respectively. Notably, in the case of the Indic dataset (Kannada), which comprises 57 writers with 228 documents, the model attained a remarkable accuracy of 100.00%, representing the highest attainable level of accuracy. Nguyen et al. [64] Utilized the JEITA-HP database of Japanese handwritten character patterns and the English databases (Firemaker and IAM) for writer identification task. In the feature extraction phase, various types of features were extracted, including local features at the region level, character level, and global features derived from multiple characters through random sampling. These features were subsequently utilized to train Convolutional Neural Networks (CNNs) for the purpose of classification, and the system obtained 99.97% identification rate on the JEITA-HP database, while delivered a commendable 91.5% identification rate on the English databases (Firemaker and IAM). Rasoulzadeh et al. [65] Implemented unified neural network architecture, featuring ResNet 20 as the primary feature extractor along with an integrated NetVLAD layer, inspired by the concept of locally aggregated descriptors (VLAD). Once the architecture was defined, it proceeded to learn direct embedding for distinct input image patches by employing the triplet semi-hard loss function. Subsequently, the embedded descriptors of every handwritten image were combined using the generalized max-pooling technique. To enhance the identification and retrieval tasks, an approach was introduced, involving the identification of the k reciprocal nearest neighbors. The experimental assessment was carried out using three publicly available datasets: CVL, ICDAR 2013, and KHATT. The outcomes were highly remarkable, boasting an accuracy of 97.41% for ICDAR, 98.6% for CVL, and 98.0% for KHATT. Wang et al. [66] Reported a U-Net model was used for the binarization process, ResNet-50 for feature extraction, and incorporates an enhanced learnable residual encoding layer to generate global descriptors. The model was put to test using the ICDAR17 competition dataset, specifically tailored for historical document writer identification, known as Historical-WI. The results indicate an impressive 72.4% accuracy rate in author identification. Table 3 gives the description about the work done in the field of different Deep learning techniques.

Table 3: Writer Identification Work in Deep Learning Techniques.

Author	Script	Dataset	Feature extraction Method	Classifier	Accuracy
Christlein et al.[34]	Medieval Manuscripts And Numerous	ICDAR 2017-Historical-WI, ICFHR16	SIFT, Raw pixels by ResNet	Exemplar SVM	76.2% 84.1%
Christlein et al.[35]	Born digital images, real scene images and real scene videos in ICDAR13 German and English Text in CVL.	ICDAR 2013 CVL	Raw pixels by CNN	Exemplar SVM	88.6% 97.8%
Chen et al.[36]	Born digital images, real scene images and real scene videos in ICDAR13.GermanandEnglish Text in CVL.	ICDAR 2013 CVL	Raw pixels by ResNet	CNN	91.8% 98%
Christlein et al.[37]	Born digital images, real scene images and real scene videos in ICDAR13.GermanandEnglish Text in CVL Arabic in KHATT	ICDAR13, CVL, KHATT	CNN activation features	Exemplar SVM	93.2%, 98.4% 98.0%
Tang et al.[38]	Born digital images, real scene images and real scene video sin ICDAR13andGerman, English Text in CVL	ICDAR 2013 CVL	Raw Pixel Using CNN	Joint Bayesian	99.1% 93%
Rehman et al.[39]	Arabic and English	QUWI	AlexNet architecture of CNN	SVM	88.11%
Yang et al.[40]	Chinese	CASIA-OLHWDB1.0 dataset	Path-signature feature and DropStroke	DCNN	99.52%.
Xing et al.[45]	English and Chinese	IAM and HWDB	Discriminative features	CNN	99.01% 93.85%.
Fiel et al. [42]	Born digital images, real scene images and real	ICDAR 2011	Raw Pixel by CNN	Nearest Neighbor	98.6% 98.3%

	scene videos in ICDAR13,Borndigital images and real scene in ICDAR2011and German, English Text in CVL	ICDAR 2013 CVL			98.3%
Nasuno et al. [60]	Japanese	(100words) from 100 writers	Raw pixels by Alenxnet	Alenxnet	90%
Kumar et al. [62]	IAM, CVL, IFN/ENIT, Kannada, and Devanagari,	Arabic-English English, Arabic, Kannada, Devanagari	Feature-level representation by region selection	DCNN	92.79 %, 99.35 %, 98.30 %, 100.0 0%, 87.06%
Nguyen et al.[64]	Japanese English	JEITA-HP (Firemaker and IAM).	Local features and global feature	CNN	91.5% 99.97%
Rasoulzadeh et al. [65]	Multi script English Arabic	ICDAR CVL KHATT	VLAD	ResNet 20	97.41 % 98.6% 98.0 %
Wang et al. [66]	(Historical-WI)	ICDAR17	Global descriptors	ResNet-50	72.4%

Additionally, the previously provided table's graphical analysis is shown in figure 10.

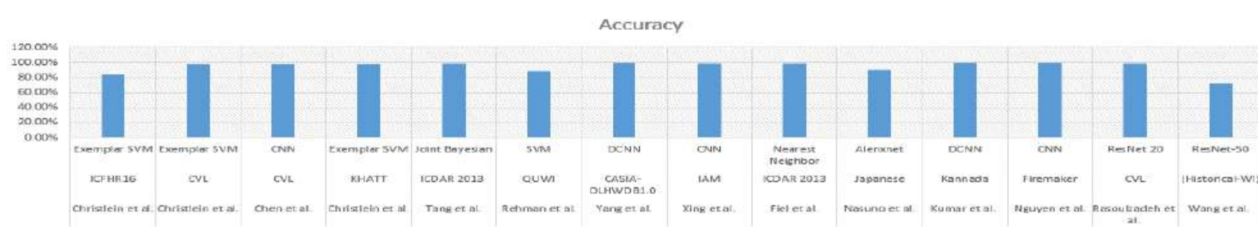

Fig 10: Graphically analysis of various dataset

6. DATASET AND SOURCE

Data-set is crucial for any research work. Having access to a dataset is crucial for making progress and evaluating research in any field, which include writer recognition and handwriting. This section will go over various datasets used for training and evaluating features for writer identification, image classification, and so on. These datasets are collected for writer identification of online and offline datasets. Table 4 discusses various datasets used for digital image processing along with source link. Several datasets have been developed for different languages in the upcoming sections.

Table 4: Description of dataset and the Source

Author Name	Year	Highlights of the article	Dataset name	Source
Singh et al. [67]	2000	The English dataset was built by the University of Buffalo's CEDAR (Center of Excellence for Document Analysis and Recognition).	CEDAR dataset contains grayscale and binary pictures of the textual content of 1,000 writers.	https://cedar.buffalo.edu/Databases

Martiet al. [68]	2002	The offline English dataset was developed by University of Bern.	IAM dataset contains 14 text lines per writer and 13,353 classified textual lines of varying content in total.	https://fki.tic.heia-fr.ch/databases/iam-handwriting-database
Liwicki et al.[69]	2005	The online English database is open to the public and can be freely used for non-commercial research.	IAM-On DB consists of over 86 K word counts from an 11 K word dictionary developed by more than 200 authors.	https://fki.tic.heia-fr.ch/databases/download-the-iam-on-line-handwriting-database
Indermühle et al.[70]	2010	The online database consists of handwritten documents that include text, diagrams, drawings, tables, formulas, lists, along with markings.	The IAM On-DO database contains 1000 documents created by around 200 writers.	https://fki.tic.heia-fr.ch/databases/download-the-iam-online-document-database
Fischer et al.[71]	2010	The dataset includes images of handwritten historical manuscripts along with corresponding ground truth data.	IAM-HistDB currently features the Saint Gall Database of the ninth century, which consists of manuscripts penned by a sole writer in Carolingian script.	https://fki.tic.heia-fr.ch/databases/iam-historical-document-database
Mahmoud [72]	2014	An open Arabic offline handwritten text database developed by KHATT Foundation(Centre for Arab typography) is an organization based in the Netherlands.	KHATT consists of 1000 handwritten forms produced by 1000 unique authors from various nations which includes paragraph, images and line images.	http://khatt.ideas2serve.net/KHATTDownload.php
Al-Ma'adeed[73]	2002	The AHDB (Arabic Handwritten Database) is a dataset specifically designed for Arabic handwritten text recognition tasks.	AHDB incorporates the most famous written Arabic phrases and textual content written by 150 writers which includes around 10,000 phrases for processing of Arabic cheque.	https://www.kaggle.com/datasets/mloey1/ahdd1
El Abed et al. [74]	2007	The Institute of Communications Technology (IFN) and the EcoleNationale des Ingenieurs de Tunis generated this dataset (ENIT).	IFN/ENIT database is organized into 26,459 images and more than 210,000 characters.	http://www.ifnenit.com/download.htm
Hussein et al.	2014	The goal is to use a modest number of basic	AlexU-Word dataset contains 25114	http://www.eng.alexu.edu.eg/~mehussein/alexu-

[75]		phrases to create a big collection of isolated Arabic words that covers all letters of the alphabet in all possible forms..	samples from 109 unique Arabic words.	word/
Strassel [76]	2009	The dataset contains audio and corresponding transcripts in Arabic, along with metadata information such as speaker IDs, language labels, and timestamps. The audio recordings cover a wide range of topics and come from various sources, including broadcast news, broadcast conversations, and web data.	The MADCAT (Multilingual Audio-Visual and Text) Arabic dataset is a large-scale multilingual dataset designed for various speech and language processing tasks which include automatic speech recognition (ASR), speaker diarization, language identification, and more.	https://www.openslr.org/48/
Kharma [77]	1999	The samples contain digits, words, signatures as well as sentences of 500 students at Jordan.	Al-Isradataset was built by 500 contributors. 37,000 words, 10,000 numbers, 2500 signatures and 500Arabic phrases with the highest popularity.	NA
Schomaker [78]	2000	Firemaker image collection for image-based pattern recognition benchmarking in forensic writer identification.	THE FIREMAKER DATASET WAS CREATED USING THE HANDWRITINGS OF 252 DUTCH STUDENTS ON 1008 SCANNED PAGES.	https://zenodo.org/record/1194612
Augustin et al.[79]	2006	A unique dataset for author identification is the Recon- naissance et Indexation de donnesManuscrites et de fac similes.	RIMES dataset includes secondary databases of characters, handwritten words, and logos totaling 300,000 snippets, as well as more than 1300 writers who have finished 5 letters, totaling 5600 letters in more than 12,000 pages with annotations.	http://www.a2ialab.com/doku.php?id=rimes_database:data:icdar2011:line:icdar2011competitionline
Freitas et al.[80]	2008	Brazilian forensic professionals and the Brazilian Federal Police in writer	The BFL database has 315 writers (Undergraduate students), each with three samples,	https://web.inf.ufpr.br/vri/databases/brazilian-forensic-letter-database/

		verification/identification is the rationale for establishing the database	totaling 945 images.	
Suet al.[81]	2006	HIT-MW handwritten writings, which are mostly authored by college students, have a balanced gender and department distribution.	In an unrestricted setting without preprinted character boxes, over 780 participants wrote 186,444 characters over 853 forms for HIT-MW.	https://code.google.com/archive/p/hit-mw-database/
Alaeiet al.[82]	2011	The Kannada Handwritten Text Database (KHTD) was an unconstrained dataset which consists of 204 handwritten documents from four different categories came from 51 native Kannada speakers.	The KHTD dataset has a total of 4298 text lines and 26115 words, respectively.	NA
Kumar et al. [83]	2013	CPAR-2012 represented a benchmark dataset for Devanagari document recognition.	The dataset includes 35,000 isolated hand written numerals, 83,300 characters, 2000 constrained and 2000 unconstrained handwritten pangrams.	NA
Bhattacharyaet al.[84]	2005	The dataset was collected by gathering numbers from postal mails as well as job application forms in Bangla and Devanagari scripts.	22,556 Devanagari numerals collected from 368 postal mails and 274 job application forms. 23,392 Bangla numerals gathered from 268 job applications and 465 mails.	https://www.isical.ac.in/~ujjwal/download/BanglaNumeral.html
Thadchanamoorthy et al.[85]	2013	Handwritten city names in Tamil, a widely used script in Sri Lanka and India.	The database contains 100 occurrences of each city name collected from 500 writers.	NA
Wilkinsonet al.[86]	1992	The National Institute of Standards and Technology, NIST, developed a databases	NIST contains a collection of handwritten forms from 3600 individuals, consisting of 810,000 isolated character images accompanied by accurate information.	https://www.nist.gov/srd/nist-special-database-19
Hull [87]	1994	A Dataset contains machine-printed Japanese character	CDER-CRDRM2	https://cedar.buffalo.edu/Databases/

		pictures.		
Hull [87]	1994	A database of handwritten words, ZIP Codes, Digits and Alphabetic characters.	CDER-CRDRM1	https://cedar.buffalo.edu/Databases/
Viard-Gaudin et al. [88]	1999	The Institute de Recherche et d'EnseignementSupérieur aux Techniques de l'Electronique, IRESTE, developed a dual on/off database.	The IRONOFF dataset included digits, characters, and words from various French writer.	http://www.iapr-tc11.org/mediawiki/index.php/Datasets_(old_page)
Serrano et al.[89]	2010	RODRIGO has a collection of various historical manuscripts written in Spanish.	The RODRIGO database encompasses across 853 pages and is organized into 307 chapters that delve into the chronicles of Spanish history.	https://zenodo.org/record/1490009
Kleber et al.[90]	2013	An Off-line Database for Writer Retrieval, Writer Identification and Word Spotting	It contains handwritten text from 311 different authors in both German and English (1 in German and 6 in English).	https://cvl.tuwien.ac.at/research/cvl-databases/an-off-line-database-for-writer-retrieval-writer-identification-and-word-spotting/
Al Maadeed et al.[91]	2012	It is a bi-script dataset.	Handwritten text in Arabic and English scripts gathered from 1017 individuals.	NA
Wang et al.[92-93]	2009-2011	CASIA: Database containing online and offline Chinese handwriting	Chinese handwritten data including paragraphs and characters collected from 1020 writers who used a digital pen to write on paper.	http://www.nlpr.ia.ac.cn/databases/handwriting/Download.html

7. ORGANIZATION OF COMPETITIONS

There has been a noticeable increase in the past few years in the number of international competitions held for various writer identification tasks. These competitions are primarily promoted on two major platforms, the International Conference on Document Analysis and Recognition (ICDAR) and the International Conference on Frontiers in Handwriting Recognition (ICFHR), which are held in conjunction with reputable document recognition conferences. Participants in these challenges are given access to training and validation datasets, and they are required to submit the executables of the algorithms they have developed, or the outcomes of their tests on the unlabeled test datasets. These competitions mostly consist of handwriting identification and associated challenges. The examination is typically performed on well-known, public handwriting databases. Despite the high level of participation, only a small fraction of groups chose to disclose their identities and share details about their algorithms. Apart from the contests mentioned in Table 5, there have been several other competitions focused on handwritten databases. However, these competitions used non-published and personal databases, which are not within the scope of our discussion.

Table 5: Organization of Competitions, participants, and Methods

Database/Script	Writers	Competition	Participants	Best Method and accuracy
(English, French, German and Greek)	26 writers	ICDAR 2011 Writer Identification Contest [94]	7 teams	TSINGHUA 100%
English and Greek (Latin based language)	250 writers	ICDAR 2013 Competition on writer Identification [95]	6 teams	CSUMD-a 99%
QUWI Database	1017	Gender Classification Competition and ICDAR 2015 Multiscript Writer Identification [96]	8 teams, 13 systems	the Nuremberg method 55%
(Historical-WI)	720 Writers	ICDAR2017 Competition on Historical Document Writer Identification (Historical-WI)[97]	5 teams	*T´ebessa II 86.6%*
Arabic dataset	54 Writers	The ICDAR2011 Arabic Writer Identification Contest [98]	30 teams	Andrew Newell and Lewis Griffin from UCL 100%
Latin script (Greek and English)	100 writers	ICFHR2012 Competition on Writer Identification Challenge 1: Latin/Greek Documents [99]	4 teams	TEBESSA-c 99%
Arabic Dataset,	206 writers	ICFHR2012 Competition on Writer Identification - Challenge 2: Arabic Scripts [100]	43 teams	Wayne Zhang and Newell and Griffin 80%
KHATT/AHTID-MW	1000 writers/23 Writer	ICFHR 2014 Arabic Writer Identification Contest [101]	3 systems	GMMS system 73.4%
QUWI	1017	ICFHR2016 Competition on Multi-script Writer Demographics Classification Using "QUWI" Database [102]	5 teams	Nuremberg Methods 84.37%

8. CONCLUSION AND FUTURE SCOPE

One type of biometric recognition that allows one for identifying the writer based on the handwritten text is writer identification system. We offer a thorough analysis of published works for both non-Indic as well as Indic scripts, together with the classifiers employed, feature extraction techniques, and accuracy levels obtained. It provides information on how dataset size can enhance a handwriting recognition system's accuracy. According to the report, there is less research on writer identification using handwriting in Indic scripts than in non-Indic scripts. The article included a number of excellent results with respectable accuracy for both Indic scripts, including Tamil, Bengali, Devanagari, Telugu, Gurmukhi and Oriya scripts, and non-Indic scripts, including Chinese, Arabic, French, Latin, Japanese, as well as Thai. The primary issue is that different Indic scripts lack a uniform database. The issue of keeping datasets with enough writers is one of the field's future prospects. The researchers' main focus will be on the advancement of novel feature extraction techniques with higher degrees of accuracy and maturity.

REFERENCES

1. Kalra K. and Rani S. (2017) Writer Identification from Offline Isolated Handwritten Gurumukhi Characters. Advances in Computational Sciences and Technology. *10*(5), 903-914. https://www.ripublication.com/acst17/acstv10n5_24.pdf.

2. Kumar M, Jindal MK, & Sharma, RK (2016). A novel framework for grading of writers using offline Gurmukhi characters. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, *86*, 405-415.https://link.springer.com/article/10.1007/s40010-016-0277-x.

3. Verma K, & Sharma, RK (2018). An efficient writing-zone identification technique for online handwritten Gurmukhi character recognition. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, *88*, 297-307. https://link.springer.com/article/10.1007/s40010-017-0416-z.

4. Sakshi, Garg NK, & Kumar M (2018). Writer identification system for handwritten Gurmukhi characters: Study of different feature-classifier combinations. In *Proceedings of International Conference on Computational Intelligence and Data Engineering: ICCIDE 2017* (pp. 125-131). Springer Singapore. https://link.springer.com/chapter/10.1007/978-981-10-6319-0_11.

5. Singh H, Sharma RK, & Singh, V P (2019). Efficient zone identification approach for the recognition of online handwritten Gurmukhi script. *Neural Computing and Applications*, *31*, 3957-3968. https://link.springer.com/article/10.1007/s00521-017-3340-x.

6. Kumar M, Jindal MK, Sharma, RK, & Jindal SR (2018). A novel framework for writer identification based on pre-segmented Gurmukhi characters. *Sādhanā*, *43*, 1-9.https://link.springer.com/article/10.1007/s12046-018-0966-z.

7. Sagar P, & Pandey A (2015). An intelligent system for writer identification of Devanagari handwritten documents using ANN. *Int J Res ApplSciEngTechnol (IJRASET)*, *3*(12), 194-201.

8. Kumar R, Kumar A, & Ahmed, P (2013). A benchmark dataset for devnagari document recognition research. *Recent Advances in Electrical Engineering Series*, (10). https://www.semanticscholar.org/paper/A-Benchmark-Dataset-for-Devnagari-Document-Research-Kumar-Kumar/5d5d4504fea88e2efb2eaff3c6ebef35d07bf70b.

9. Halder C, Thakur K, Phadikar S, & Roy K (2015). Writer identification from handwritten Devanagari script. In *Information Systems Design and Intelligent Applications: Proceedings of Second International Conference INDIA 2015, Volume 2* (pp. 497-505). Springer India. https://www.researchgate.net/publication/268814031_Writer_Identification_from_Handwritten_Devanagari_Script.

10. Choksi AH, & Thakkar SP (2012). Recognition of similar appearing Gujarati characters using fuzzy-knn algorithm. *International Journal of Computer Applications*, *55*(6). https://research.ijcaonline.org/volume55/number6/pxc3882666.pdf.

11. Desai AA (2015). Support vector machine for identification of handwritten Gujarati alphabets using hybrid feature space. *CSI transactions on ICT*, *2*(4), 235-241. https://link.springer.com/article/10.1007/s40012-014-0059-z.

12. Rahiman MA, &Rajasree MS (2009). Printed Malayalam character recognition using back-propagation neural networks. In *2009 IEEE International Advance Computing Conference* (pp. 197-201). IEEE. DOI: 10.1109/IADCC.2009.4809006.

13. Idicula SM (2011). A novel approach to writer identification in Malayalam using graphemes. In *Computer Networks and Intelligent Computing: 5th International Conference on Information Processing, ICIP 2011, Bangalore, India, August 5-7, 2011. Proceedings* (pp. 646-651). Springer Berlin Heidelberg. https://link.springer.com/chapter/10.1007/978-3-642-22786-8_81.

14. Manjusha K, Kumar MA, &Soman KP (2019). On developing handwritten character image database for Malayalam language script. *Engineering Science and Technology, an International Journal*, *22*(2), 637-645. https://doi.org/10.1016/j.jestch.2018.10.011.

15. Chanda, S, Franke K, & Pal U (2012). Text independent writer identification for Oriya script. In *2012 10th IAPR International Workshop on Document Analysis Systems* (pp. 369-373). IEEE. DOI:10.1109/DAS.2012.86.

16. Mishra, T. K. (2015). *Development of features for recognition of handwritten Odia characters* (Doctoral dissertation). https://core.ac.uk/download/pdf/80147738.pdf.

17. Panda SR, &Tripathy J (2015, September). Odia offline typewritten character recognition using template matching with unicode mapping. In *2015 international symposium on advanced computing and communication (ISACC)* (pp. 109-115). IEEE. DOI: 10.1109/ISACC.2015.7377325.

18. Jayanthi SK, &Rajalakshmi D (2011). Writer identification for offline Tamil handwriting based on gray-level co-occurrence matrices. In *2011 Third International Conference on Advanced Computing* (pp. 187-192). IEEE. DOI: 10.1109/ICoAC.2011.6165173.

19. Thendral T, Vijaya MS, &Karpagavalli S (2013). Supervised learning approach for Tamil writer identity prediction using global and local features. *Int J Res EngTechnol (IJRET)*, *2*(5), 204-208.

20. Thendral T, Vijaya MS, &Karpagavalli S (2015). Prediction of writer using Tamil handwritten document image based on pooled features. *International Journal of Computer and Information Engineering*, *9*(6), 1572-1578. https://zenodo.org/record/1108955.

21. Thendral T, Vijaya MS, &Karpagavalli S (2014). Analysis of Tamil character writings and identification of writer using Support Vector Machine. In *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies* (pp. 1407-1411). IEEE. DOI: 10.1109/ICACCCT.2014.7019332.

22. Purkait P, Kumar R, &Chanda B (2010). Writer identification for handwritten Telugu documents using directional morphological features. In *2010 12th International Conference on Frontiers in Handwriting Recognition* (pp. 658-663). IEEE. DOI: 10.1109/ICFHR.2010.108.

23. Andrew C, Reddy S, Pulabaigari V, & Pal U (2017). Text independent writer identification for Telugu script using directional filter based features. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (Vol. 5, pp. 65-70). IEEE. DOI: 10.1109/ICDAR.2017.330.

24. Mukarambi, G., &Hangarge, M. (2012). Writer Identification By Texture Analysis Based On Kannada Handwriting.https://www.academia.edu/2845316/Writer_Identification_By_Texture_Analysis_Based_On_Kannada_Handwriting.

25. Asi A, Abdalhaleem A, Fecker D, Märgner V, & El-Sana J (2017). On writer identification for Arabic historical manuscripts. *International Journal on Document Analysis and Recognition (IJDAR)*, *20*, 173-187. https://link.springer.com/article/10.1007/s10032-017-0289-3.

26. Chaabouni A, Boubaker H, Kherallah M, El Abed H, &Alimi AM (2014). Static and dynamic features for writer identification based on multi-fractals. *Int. Arab J. Inf. Technol.*, *11*(4), 416-424. https://iajit.org/PDF/vol.11,no.4/5240.pdf.

27. Tan GX, Viard-Gaudin C, &Kot AC (2009). Automatic writer identification framework for online handwritten documents using character prototypes. *pattern recognition*, *42*(12), 3313-3323. https://doi.org/10.1016/j.patcog.2008.12.019.

28. Shin J, & Liu Z (2015). Writer identification method using inter and intra-stroke information. *INFOCOMP 2015*, 92. https://www.academia.edu/65562767/Writer_Identification_Method_using_Inter_and_Intra_Stroke_Information.

29. Li X, & Ding X (2009). Writer identification of chinese handwriting using grid microstructure feature. In *Advances in Biometrics: Third International Conference, ICB 2009, Alghero, Italy, June 2-5, 2009. Proceedings 3* (pp. 1230-1239). Springer Berlin Heidelberg. https://link.springer.com/chapter/10.1007/978-3-642-01793-3_124.

30. Bertolini D, Oliveira LS, Justino, E, &Sabourin R (2013). Texture-based descriptors for writer identification and verification. *Expert Systems with Applications*, *40*(6), 2069-2080. https://doi.org/10.1016/j.eswa.2012.10.016.

31. Fiel S, &Sablatnig R (2012). Writer retrieval and writer identification using local features. In *2012 10th IAPR International Workshop on Document Analysis Systems* (pp. 145-149). IEEE. DOI: 10.1109/DAS.2012.99

32. Siddiqi I, & Vincent N (2007, September). Writer identification in handwritten documents. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)* (Vol. 1, pp. 108-112). IEEE. DOI: 10.1109/ICDAR.2007.4378685.

33. Sreeraj M, &Idicula SM (2011). A survey on writer identification schemes. *International Journal of Computer Applications*, *26*(2), 23-33. https://www.ijcaonline.org/volume26/number2/pxc3874205.pdf.

34. Christlein V, Gropp M, Fiel, S, & Maier A (2017, November). Unsupervised feature learning for writer identification and writer retrieval. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (Vol. 1, pp. 991-997). IEEE. https://arxiv.org/pdf/1705.09369.pdf.

35. Christlein V, Bernecker D, Maier A, &Angelopoulou E (2015). Offline writer identification using convolutional neural network activation features. In *Pattern Recognition: 37th German Conference, GCPR 2015, Aachen, Germany, October 7-10, 2015, Proceedings 37* (pp. 540-552). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-319-24947-6_45.

36. Chen S, Wang Y, Lin CT, Ding W, & Cao Z. (2019). Semi-supervised feature learning for improving writer identification. *Information Sciences*, *482*, 156-170. https://doi.org/10.1016/j.ins.2019.01.024.

37. Christlein V, & Maier A (2018, April). Encoding CNN activations for writer recognition. In *2018 13th IAPR international workshop on document analysis systems (DAS)* (pp. 169-174). IEEE. DOI:10.1109/DAS.2018.9.

38. Tang Y, & Wu X (2016, October). Text-independent writer identification via CNN features and joint Bayesian. In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (pp. 566-571). IEEE. DOI**:** 10.1109/ICFHR.2016.0109.

39. Rehman A, Naz S, Razzak MI, & Hameed, I. A. (2019). Automatic visual features for writer identification: a deep learning approach. *IEEE access*, *7*, 17149-17157. DOI: 10.1109/ACCESS.2018.2890810.

40. Yang W, Jin L, & Liu M(2015, August). Chinese character-level writer identification using path signature feature, DropStroke and deep CNN. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (pp. 546-550). IEEE. DOI:10.1109/ICDAR.2015.7333821.

41. Zhu Y, & Wang Y (2016, December). An offline text-independent writer identification system with SAE feature extraction. In *2016 International Conference on Progress in Informatics and Computing (PIC)* (pp. 432-436). IEEE. DOI:10.1109/PIC.2016.7949540.

42. Fiel S, &SablatnigR (2015). Writer identification and retrieval using a convolutional neural network. In *Computer Analysis of Images and Patterns: 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015, Proceedings, Part II 16* (pp. 26-37). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-319-23117-4_3.

43. Pandey P, &Seeja KR (2018). Forensic writer identification with projection profile representation of graphemes. In *Proceedings of First International Conference on Smart System, Innovations and Computing: SSIC 2017, Jaipur, India* (pp. 129-136). Springer Singapore. https://link.springer.com/chapter/10.1007/978-981-10-5828-8_13.

44. Bertolini D, Oliveira LS, &Sabourin R (2016, December). Multi-script writer identification using dissimilarity. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 3025-3030). IEEE. DOI: 10.1109/ICPR.2016.7900098.

45. Xing L, &Qiao Y (2016, October). Deepwriter: A multi-stream deep CNN for text-independent writer identification. In *2016 15th international conference on frontiers in handwriting recognition (ICFHR)* (pp. 584-589). IEEE. DOI: 10.1109/ICPR.2016.7900098.

46. Durou A, Aref I, Al-Maadeed S, Bouridane A, &BenkhelifaE(2019). Writer identification approach based on bag of words with OBI features. *Information Processing & Management*, *56*(2), 354-366. https://www.sciencedirect.com/science/article/pii/S0306457316305829.

47. Venugopal V, &Sundaram S (2017). An online writer identification system using regression-based feature normalization and codebook descriptors. *Expert Systems with Applications*, *72*, 196-206. https://www.sciencedirect.com/science/article/pii/S095741741630673X.

48. Wu, Y., Lu, H., & Zhang, Z. (2017). Text-independent online writer identification using hidden markov models. *IEICE TRANSACTIONS on Information and Systems*, *100*(2), 332-339. DOI:10.1587/transinf.2016EDP7238.

49. Garz A, Würsch M, Fischer A, &Ingold R (2016). Simple and fast geometrical descriptors for writer identification. *Electronic Imaging*, *2016*(17), 1-12. file:///C:/Users/Sukh/Downloads/EI_2016_art00005_Angelika-Garz.pdf.

50. Wu X, Tang Y, & Bu W (2014). Offline text-independent writer identification based on scale invariant feature transform. *IEEE Transactions on Information Forensics and Security*, *9*(3), 526-536. DOI:10.1109/TIFS.2014.2301274.

51. Hu Y, Yang W, & Chen Y (2014, October). Bag of features approach for offline text-independent Chinese writer identification. In *2014 IEEE International Conference on Image Processing (ICIP)* (pp. 2609-2613). IEEE. https://projet.liris.cnrs.fr/imagine/pub/proceedings/ICIP-2014/Papers/1569902795.pdf.

52. Fecker, D., Asit, A., Märgner, V., El-Sana, J., &Fingscheidt, T. (2014, August). Writer identification for historical Arabic documents. In *2014 22nd International conference on pattern recognition* (pp. 3050-3055). IEEE. DOI:10.1109/ICPR.2014.526.

53. Christlein V, Bernecker D, Hönig F, &Angelopoulou E (2014, March). Writer identification and verification using GMM supervectors. In *IEEE Winter Conference on Applications of Computer Vision* (pp. 998-1005). IEEE. DOI:10.1109/WACV.2014.6835995.

54. Newell AJ, & Griffin LD (2014). Writer identification using oriented basic image features and the delta encoding. *Pattern Recognition*, *47*(6), 2255-2265. https://doi.org/10.1016/j.patcog.2013.11.029.

55. He ZY, & Tang YY (2004, August). Chinese handwriting-based writer identification by texture analysis. In *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 04EX826)* (Vol. 6, pp. 3488-3491). IEEE. DOI: 10.1109/ICMLC.2004.1380392.

56. Khalifa E, Al-Maadeed S, Tahir MA, Bouridane A, &Jamshed A (2015). Off-line writer identification using an ensemble of grapheme codebook features. *Pattern Recognition Letters*, *59*, 18-25. https://doi.org/10.1016/j.patrec.2015.03.004.

57. Christlein V, Bernecker D, Hönig F, &Angelopoulou E. (2016). Writer Identification Using GMM Supervectors and Exemplar-SVMs. Pattern Recognition. DOI:10.1016/j.patcog.2016.10.005.

58. Hannad Y, Siddiqi I, & El Kettani MEY (2016). Writer identification using texture descriptors of handwritten fragments. *Expert Systems with Applications*, *47*, 14-22. DOI:10.1016/j.eswa.2015.11.002.

59. Chahi A, Ruichek Y, &Touahni R (2018). Block wise local binary count for off-line text-independent writer identification. *Expert Systems with Applications*, *93*, 1-14. https://doi.org/10.1016/j.eswa.2017.10.010.

60. Nasuno R, & Arai S (2017, September). Writer identification for offline japanese handwritten character using convolutional neural network. In *Proceedings of the 5th IIAE (Institute of Industrial Applications Engineers) International Conference on Intelligent Systems and Image Processing* (pp. 94-97). https://pdfs.semanticscholar.org/6660/a0754e644cef63aa54210c8241516998aa52.pdf.

61. Chahi A, Ruichek Y, &Touahni R (2020). Local gradient full-scale transform patterns based off-line text-independent writer identification. *Applied Soft Computing*, *92*, 106277. https://doi.org/10.1016/j.asoc.2020.106277.

62. Kumar P, & Sharma A (2020). Segmentation-free writer identification based on convolutional neural network. *Computers & Electrical Engineering*, *85*, 106707. https://doi.org/10.1016/j.compeleceng.2020.106707.

63. Pandey S, Harit G (2017). Writer Identification for Handwritten Words. In: ,*et al.* Computer Vision, Graphics, and Image Processing. ICVGIP 2016. Lecture Notes in Computer Science, vol 10481. Springer, Cham. https://doi.org/10.1007/978-3-319-68124-5_23.

64. Nguyen HT, Nguyen CT, Ino T, Indurkhya B, & Nakagawa M (2019). Text-independent writer identification using convolutional neural network. *Pattern Recognition Letters*, *121*, 104-112. https://arxiv.org/ftp/arxiv/papers/2009/2009.04877.pdf.

65. Rasoulzadeh S, &BabaAli B (2022). Writer identification and writer retrieval based on NetVLAD with Re-ranking. *IET Biometrics*, *11*(1), 10-22. https://doi.org/10.1049/bme2.12039.

66. Wang Z, Maier A, &Christlein V (2021). Towards end-to-end deep learning-based writer identification. *INFORMATIK 2020*. https://dl.gi.de/server/api/core/bitstreams/5ed98c15-7f44-4d6d-9b80-832956f4d389/content.

67. Singh S, &Hewitt M (2000, September). Cursive digit and character recognition in CEDAR database. In *Proceedings 15th international conference on pattern recognition. ICPR-2000* (Vol. 2, pp. 569-572). IEEE. DOI**:** 10.1109/ICPR.2000.906138.

68. Marti UV, &Bunke H (2002). The IAM-database: an English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, *5*, 39-46. https://link.springer.com/article/10.1007/s100320200071.

69. Liwicki, M., &Bunke, H. (2005, August). IAM-OnDB-an on-line English sentence database acquired from handwritten text on a whiteboard. In *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)* (pp. 956-961). IEEE**.** DOI: 10.1109/ICDAR.2005.132.

70. Indermühle E, Liwicki M, &Bunke H (2010, June). IAMonDo-database: an online handwritten document database with non-uniform contents. In *Proceedings of the 9th IAPR international workshop on document analysis systems* (pp. 97-104 https://fki.tic.heia-fr.ch/static/pdf/das10db.pdf.

71. Fischer A, Indermühle E, Bunke H, Viehhauser G, &Stolz M (2010, June). Ground truth creation for handwriting recognition in historical documents. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems* (pp. 3-10). https://fki.tic.heia-fr.ch/static/pdf/fischer10groundtruth-preprint.pdf.

72. Mahmoud SA, Ahmad I, Al-Khatib WG., Alshayeb M, Parvez MT, Märgner V, & Fink G A (2014). KHATT: An open Arabic offline handwritten text database. *Pattern Recognition*, *47*(3), 1096-1112. https://doi.org/10.1016/j.patcog.2013.08.009.

73. Al-Ma'adeed S, Elliman D, & Higgins C A (2002, August). A data base for Arabic handwritten text recognition research. In *Proceedings eighth international workshop on frontiers in handwriting recognition* (pp. 485-489). IEEE. https://www.kaggle.com/datasets/mloey1/ahdd1.

74. El Abed H, &Margner V (2007, February). The IFN/ENIT-database-a tool to develop Arabic handwriting recognition systems. In *2007 9th International Symposium on Signal Processing and Its Applications* (pp. 1-4). IEEE. http://www.ifnenit.com/download.htm .

75. Hussein M E, Torki M, Elsallamy A., &Fayyaz M (2014). Alexu-word: a new dataset for isolated-word closed-vocabulary offline arabic handwriting recognition. *arXiv preprint arXiv:1411.4670*. http://www.eng.alexu.edu.eg/~mehussein/alexu-word/

76. Strassel S (2009, April). Linguistic resources for Arabic handwriting recognition. In *Proceedings of the Second International Conference on Arabic Language Resources and Tools, Cairo, Egypt*. https://www.openslr.org/48/

77. Kharma N, Ahmed M, & Ward R (1999, May). A new comprehensive database of handwritten Arabic words, numbers, and signatures used for OCR testing. In *Engineering Solutions for the Next Millennium. 1999 IEEE Canadian Conference on Electrical and Computer Engineering (Cat. No. 99TH8411)* (Vol. 2, pp. 766-768). IEEE.DOI: 10.1109/CCECE.1999.808042

78. Schomaker L, Vuurpijl L, &Schomaker L (2000). Forensic writer identification: A benchmark data set and a comparison of two systems.https://research.rug.nl/en/publications/forensic-writer-identification-a-benchmark-data-set-and-a-compari.

79. Augustin E, Carré M, Grosicki E, Brodin JM, Geoffrois E, &Prêteux F (2006). RIMES evaluation campaign for handwritten mail processing. In *International Workshop on Frontiers in Handwriting Recognition (IWFHR'06),* (pp. 231-235). DOI: 10.1109/ICDAR.2009.224.

80. Freitas C, Oliveira LS, Sabourin R, &Bortolozzi F (2008). Brazilian forensic letter database. In *11th International workshop on frontiers on handwriting recognition, Montreal, Canada*. https://www.researchgate.net/publication/228813177_Brazilian_forensic_letter_database.

81. Su T, Zhang T, & Guan D (2006, October). HIT-MW dataset for offline Chinese handwritten text recognition. In *Tenth International Workshop on Frontiers in Handwriting Recognition*. Suvisoft. https://www.researchgate.net/publication/228958295_HIT-MW_Dataset_for_Offline_Chinese_Handwritten_Text_Recognition.

82. Alaei A, Nagabhushan P, & Pal U (2011, September). A benchmark Kannada handwritten document dataset and its segmentation. In *2011 International Conference on Document Analysis and Recognition* (pp. 141-145). IEEE. DOI: 10.1109/ICDAR.2011.37.

83. Kumar R, Kumar A, & Ahmed P (2013). A benchmark dataset for devnagari document recognition research. *Recent Advances in Electrical Engineering Series*, (10). https://www.semanticscholar.org/paper/A-Benchmark-Dataset-for-Devnagari-Document-Research-Kumar-Kumar/5d5d4504fea88e2efb2eaff3c6ebef35d07bf70b.

84. Bhattacharya U, & Chaudhuri BB (2005, August). Databases for research on recognition of handwritten characters of Indian scripts. In *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)* (pp. 789-793). IEEE. DOI: 10.1109/ICDAR.2005.84.

85. Thadchanamoorthy S, Kodikara ND, Premaretne, HL, Pal U, & Kimura F (2013, August). Tamil handwritten city name database development and recognition for postal automation. In *2013 12th International Conference on Document Analysis and Recognition* (pp. 793-797). IEEE. DOI: 10.1109/ICDAR.2013.162

86. Wilkinson R, Geist J, Janet S, Grother P, Burges C, Creecy R, & Wilson C. (1992). The first census optical character recognition systems conf.# NISTIR 4912. *The US Bureau of Census and the National Institute of Standards and Technology, Gaithersburg, MD*. https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=905664.

87. Hull JJ (1994). A database for handwritten text recognition research. *IEEE Transactions on pattern analysis and machine intelligence*, *16*(5), 550-554. DOI: 10.1109/34.291440.

88. Viard-Gaudin C, Lallican PM, Knerr S, &Binter P (1999, September). The ireste on/off (ironoff) dual handwriting database. In *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318)* (pp. 455-458). IEEE. DOI: 10.1109/ICDAR.1999.791823

89. Serrano N, Castro F, & Juan A (2010, May). The RODRIGO Database. In *LREC* (pp. 19-21). https://www.researchgate.net/publication/220746542_The_RODRIGO_Database.

90. Kleber F, Fiel S, Diem M, &Sablatnig R (2013, August). Cvl-database: An off-line database for writer retrieval, writer identification and word spotting. In *2013 12th international conference on document analysis and recognition* (pp. 560-564). IEEE.10.1109/ICDAR.2013.117.

91. Al Maadeed S, Ayouby W, Hassaine A, &Aljaam JM (2012, September). QUWI: an Arabic and English handwriting dataset for offline writer identification. In *2012 International Conference on Frontiers in Handwriting Recognition* (pp. 746-751). IEEE. DOI: 10.1109/ICFHR.2012.256.

92. Wang DH, Liu CL., Yu JL, & Zhou, X. D. (2009, July). CASIA-OLHWDB1: A database of online handwritten Chinese characters. In *2009 10th International Conference on Document Analysis and Recognition* (pp. 1206-1210). IEEE. DOI: 10.1109/ICDAR.2009.163.

93. Liu CL, Yin F, Wang DH, & Wang QF (2011, September). CASIA online and offline Chinese handwriting databases. In *2011 international conference on document analysis and recognition* (pp. 37-41). IEEE. https://www.researchgate.net/publication/232262651_CASIA_Online_and_Offline_Chinese_Handwriting_Databases

94. Louloudis G, Stamatopoulos N, & Gatos B (2011, September). ICDAR 2011 writer identification contest. In *2011 International Conference on Document Analysis and Recognition* (pp. 1475-1479). IEEE.DOI: 10.1109/ICDAR.2011.293.

95. Louloudis G, Gatos B, Stamatopoulos N, & Papandreou A (2013, August). ICDAR 2013 competition on writer identification. In *2013 12th International Conference on Document Analysis and Recognition* (pp. 1397-1401). IEEE. DOI: 10.1109/ICDAR.2013.282.

96. Djeddi C, Al-Maadeed S, Gattal A, Siddiqi I, Souici-Meslati L, & El Abed H (2015, August). ICDAR2015 competition on multi-script writer identification and gender classification using ‘QUWI’database. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (pp. 1191-1195). IEEE.DOI: 10.1109/ICDAR.2015.7333949.

97. Fiel S, Kleber F, Diem M, Christlein V, Louloudis G, Nikos S, & Gatos B (2017, November). Icdar2017 competition on historical document writer identification (historical-wi). In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (Vol. 1, pp. 1377-1382). IEEE.DOI: 10.1109/ICDAR.2017.225.

98. Hassaïne A, Al-Maadeed S, Alja'am JM, Jaoua A, &Bouridane A (2011, September). The ICDAR2011 Arabic writer identification contest. In *2011 International Conference on Document Analysis and Recognition* (pp. 1470-1474). IEEE.DOI: 10.1109/ICDAR.2011.292.

99. Louloudis G, Gatos B, &Stamatopoulos N (2012, September). ICFHR 2012 competition on writer identification challenge 1: Latin/Greek documents. In *2012 International Conference on Frontiers in Handwriting Recognition* (pp. 829-834). IEEE.DOI: 10.1109/ICFHR.2012.219

100. Hassaïne A, & Al Maadeed S (2012, September). ICFHR 2012 competition on writer identification challenge 2: Arabic scripts. In *2012 International Conference on Frontiers in Handwriting Recognition* (pp. 835-840). IEEE. DOI**:** 10.1109/ICFHR.2012.218.

101 Slimane F, Awaida S, Mezghani A, Parvez MT, Kanoun, S, Mahmoud SA, &Märgner V (2014, September). Icfhr2014 competition on arabic writer identification using ahtid/mw and khatt databases. In *2014 14th international conference on frontiers in handwriting recognition* (pp. 797-802). IEEE.DOI: 10.1109/ICFHR.2014.139.

102. Djeddi C, Al-Maadeed S, Gattal A, Siddiqi I, Ennaji A, & El Abed H (2016, October). ICFHR2016 Competition on multi-script writer demographics classification using" QUWI" database. In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (pp. 602-606). IEEE.DOI: 10.1109/ICFHR.2016.0115.