

Academic Journal of Nawroz University (AJNU), Vol.10, No.1, 2021 This is an open access article distributed under the Creative Commons Attribution License Copyright ©2017. e-ISSN: 2520-789X https://doi.org/10.25007/ajnu.v10n1a862



# Kurdish Handwritten Classification Based on Human Personality Analysis

Ahmed Abdullah Ahmed

Faculty of Engineering & Computer Science, Qaiwan International University (QIU), Raparin, Sulaymaniyah, Kurdistan Region-Iraq

# ABSTRACT

This study proposes a holistic technique of classifying Kurdish handwritten text documents, involving several stages. The first stage entails the sectioning of Kurdish handwritten document images into lines, words and characters and the second stage entails the obtaining of structural and statistical features from the sectioned parts that are often utilized in human personality analysis for the examination of feature behavior and effectiveness. This is done by combining the entire potential outcomes to determine the significant features set. The third stage entails the use of F-Measure to evaluate the extracted features performance and their combination in various relationship methods, individually and in groups. The last stage entails the actual experiment using the standard KRDOH dataset of the Kurdish handwritten text, containing 1076 volunteers' samples of different ages, genders and education levels in a cumulative 4304 manuscripts consisting of 4 contributed pages by each writer. Based on the results obtained from multiple runs of individual clustering method of each distance measures, good set of features generally deliver significant enhanced clustering of handwritten structures.

Keywords: Grouping, writer identification, feature extraction, human personality, feature combination.

## 1. Introduction

Majority of scientific fields in engineering like intelligent retrieval, image processing, and facial recognition have pattern recognition as part of the top pertinent areas, and from pattern recognition, one of the important fields is handwriting analysis, particularly Writer Identification (WI). Throughout the past decade, writer identification of handwriting samples has been extensively examined. Various methods of writer recognition have been created and handle different proposed to handwriting scripts/languages, with each script characterized by unique and specific elements that can be examined through a distinct approach. Thus, this enables a set of writer-specific features to be obtained from writing samples volunteered by an individual and conduct a comparison through various samples, so that the writer can be identified. According to past studies [1] [2] [3] [4] [5] [6], handwriting has been deemed as an individualistic, with the individuality of the writer being directed by the hypothesis that each individual provides a consistent style of handwriting, urging researchers in the field to further understand and find the top solution for writer identification. Despite the possibilities, several challenges are still faced owing to the limited human capability of observing and recognizing various handwriting styles.

When comparing one's demeanor to the strokes of one's writing, it can be argued that the person's mental health can be discerned from his posture, happy appearance, depressed appearance, where in the latter, the person's shoulders may be slumped, and the person may talk with heavy and flat voice. In case the person is happy, his head would be up, shoulders back, face smiling and voice full of vitality. Similarly, the strokes of the pen on a piece of paper is an expression of the feelings of the writer at the time that the writing took place and this is often aligned with the person's facial expressions, voice tone and overall body language. In other words, the writer's writing style and words selection indicates his experience with the world and his behavior. However, graphologists opt to conduct an analysis of the hand movements involved in writing the script, rather than examining the hand movements (body language [7].

Thus, this work is primarily focused on sub-dividing data objects into smaller classes, in a way that each class has a high level of similarity (indicating same writer) and low level of similarity (indicating different writer), in order to determine if the author utilized the top and effective handwriting style features utilized by human personality analysis experts [7][8]. This facilitates the writer identification process in an expedient and accurate manner.

## 2. RELATED WORK

In the past several years, Latin or Chinese handwritten documents have been extensively focused on, while Kurdish handwritten documents were largely under-researched, with studies mainly aiming to obtain individual features from a handwritten sample in order to distinguish the writer. In the present work, handwritten samples classification is tackled by reviewing some techniques of writer identification and classification. The performance of pattern recognition is largely dependent on learning scheme, feature extraction, and classification [9] [10], where both selection and extraction of suitable features playing a key part in the process of pattern recognition, which is what writer identification top issue is all about (the extraction of features from different handwriting styles to determine the writer [11] [12] [13] [14] [15] [3] [16]. Then the features extracted from the writing styles are classified and placed into groups/classes that they belong under.

A person's handwriting in terms of shape and style is unique from those of others, and for that person, handwriting in different periods sometimes differ. Nevertheless, every person has his/her own individualistic writing style [1][2][3][5][17][6], with the person's many writing styles having a distinct texture [6]. In this regard, some features of the writing style can be generalized as individual features representing the writer's individuality. Literature shows that rigid characteristics like stroke, slope, length, baseline, position, height, and blackness gray distribution have been extracted but the cursive word's entire shape has not been examined. In relation to this, a general behavior in real systems reveals the decrease of accuracy with the increase of reference vectors that the lexicon has [18]. In the same way, computational process complexity stems from the lexicon that increases with the increase of its size [19]. However, the global approach excludes new lexicon into the data record [20] [21] [22]. In addition, the handwriting of a person is also presented through rigid characteristics and this contributes to the major differences among features for same writer and low differences for different ones. In the process of classification, intraclass correlation (same writer) and interclass correlation (different writers) differences are pertinent and thus, human personality analysis methods and the meaning of the features are focused intently on, just as they are in this study [23][24][25][26].

In the writer identification process, classification is vital in classifying obtained features in order to determine the writer that the features correspond to. Two major types of classes need to be considered during handwriting comparison and that is intraclass correlation and Interclass correlation. More specifically, intraclass focuses on a word/script written by the same person, notwithstanding the style of writing and shape, while interclass focuses on any word/script written by different writers. The variation in the former class has to be lower than that of the latter one [27] [28] [5] [29]. Also, handwriting style differences pose challenges for automatic transcription of handwritten papers and thus such papers are transcribed by hand [30] [31], involving a time consuming, unreliable and tiring task. Thus, to increase the task's accuracy, efficiency and time consumption, automatic word recognition is needed to segment handwritten text documents into words that can be classified based on features into clusters, with each cluster enclosing a specific word [32] [33]. For ease of access, the clusters are appropriated indices in handwritten documents. Generally speaking, the automatic handwriting recognition methods are rife with complexity [34] and, in distinct domains, the images of handwriting recognition also have their limitations [35].

According to [22], traditional handwriting recognition methods call for high features extraction accuracy along with recognition steps [22] and a recognition system generally comprises two steps [36]. The first step entails the extraction of features, where the input text data is exposed to extraction measurements. Extracting features from the input data are difficult and such difficulties are handled by obtaining the top significant information concerning the words classification to allow words distinction. The second step entails classification, defining the class to which the input word belongs under. Feature extraction is the most crucial step in the character recognition approach in order to realize a reliable rate of recognition [37]. In relation to this, there are many feature extraction methods that have been brought forward in literature that are statistical and structural methods. Specifically, statistical features are obtained from the pixels statistical distribution in the handwritten document image, while structural features are topological and geometrical features in the patterns of the handwritten text [38].

The classification of the extracted features calls for the use of a neural network or a machine learning approach, which differentiate among handwriting styles and considers new word's shapes [39]. Handwritten text application development requires careful and efficient planning of an automated handwritten recognition system stages as the problem to be faced is full of complexity and variability [36].

#### Academic Journal of Nawroz University (AJNU), Vol.10, No.1, 2021

Literature has revealed different writing styles based on different geographical locations, cultural backgrounds, ages, genders, among others [40] [41]. And as people age, their writing styles often change. Moreover, in cursive handwriting, the characteristics (e.g., ascenders/descenders height, word length, letter etc.) usually make for different concavities, handwriting styles. In the same way, letters can be linked in different ways, leading to different letter standards, where sometimes they can be illegible. Variability in this type of handwriting is due to the style of the writer and the geometric factors that are brought on by the writing conditions. In majority of off-line instances, little or no control is placed on the used instrument and its type. This challenge is compounded by the complex interactions between instrument and the later operations (i.e., scanning and binerization) in algorithms, when it comes to offline handwriting recognition. Also, low quality images (e.g., broken lines) brought on by the machine printers or fax machines, which causes tremendous challenge for the pattern recognition methods.

Over time, writing styles and forms and knowledge related to them are predicted to evolve, allowing paleographers identify the time and location within which the manuscript was written. Most of the ancient manuscripts can be found in archives, libraries and private collections and they can be useful for paleographers to conduct manuscript dating, authentication and examination of the relationship between writing styles and writers. Writing styles classification and analysis is similar to writers' recognition although the latter calls for accurate decision-making regarding the appropriation of a script to a specific class.

In the present study, the author concentrates on the primary process prior to that of writer recognition as it has a direct impact on the identification's time and precision – this process is known as handwriting classification. The study depends on significant features that are used by human personality analysts to categorize the samples into small sections known as classes, with each class representing a high level of similarity (indicating same writer), while lower level of similarity (indicating different writer). The entire features and combination potential and their relationship are examined, testing the methods utilized to obtain the distances between the handwriting samples in order to realize the study's main objective [42] [43].

### 3. DATABASE

This study uses the largest Kurdish offline handwriting database for the experiments. More specifically, Kurdish Offline Handwritten Database (KRDOH) includes 4304 handwritten forms, supplied by 1076 writers of different education levels, genders and ages. The dataset segments have been taken out of the Kurdish poetry, scanned in 200, 300 and 600 dpi resolution. The characters of the manuscript number 4304 and consist of 17,466 lines [44]. A sample from the KRDOH data base is provided in fig. 1.

خانها «دو باری مدریّ) برت ایسه ۹ وازیه ایر شاری خونه دیستژ ( طاقاتی یا داشتهه ، برکی ازاری خدانم حمایت کم می خانه ا شاته بوشگیکی جاری خودی تا رانیه . موه	لمای شاره کری کلیته می جاد و اینوی ایتری چان ترواح عامی با جاری با بیم هناد رکی زور امر سال سرتارم عام ساریک نال بالم را کری که کمای ماری روستوره می کرد م ناک بیده کمای اهراکه برونومای دوردون می کرد م ناک بیده کمای اهراکه برونومای دوردون
صادام به سرور شدیکان از توکی هستگرد نه و نتریک بود به میکدد. با او پروی در مدید سدیکا فالدی سرتونا نامکار به د کاکرمنا لام برانگرو، در نترانیک بعربرمیا روانی دیکانه ناما نبع) مده مستول کافته ک	مکرکوها و چورکاییه کا مه عدینهادا، الم تاریخه بالکالو کارکونه و مه و موجع کا تشیستار تا نوه میترا را نام هنگوه کا بورسیاوتیکا فاته دیکار نیوشنها فار کالمهمار هند دونانه که ستیمک تعملک.
Page 04	Page 03

Fig. 1. Sample Form of KRDOH Database.

## 4. METHODOLOGY

In the present study, the main aim is writer classification of offline handwritten documents to examine, determine and apply appropriate methods to extract the significant features from the texts and conduct clustering. The study model is presented in a block diagram (see fig 2). The modules are described in detail in the next sub-sections.



## 4.1 Pre-processing

In the pre-processing stage, data is gathered and prepared into suitable forms of working data to categorize them, and for images, the step entails scanning, digitization, and manual/automatic cleaning and improvement. The step is carried out using biometric system components collection and processing. In instances of offline writer identification,

### Academic Journal of Nawroz University (AJNU), Vol.10, No.1, 2021

the stage takes on the following process; text images digitization, skew correction, conversion to a binary image and text component extraction to support the subsequent processing stage. Notably, the binarization process impacts the intensity of the handwriting samples and to steer clear of such impact, the researchers conduct features extraction prior the process of binarization. The pseudo-code of binarization process is given in Algorithm 1. Binarized version is demonstrated in fig 3.

```
Algorithm 1: Pseudo-code of binarization process
```

Fig 3: Image Binarization

### 4.2 Connected Components & Punctuation Removal

This refers to a complete unit of linked ink trace that is surrounded by the points where the writer's pen has been lifted. The first phase in graphemes extraction is selecting the whole connected components from the input page, and each Connected Component CC, is described through its bounding box (x1 (CC), y1 (CC), x2 (CC), y2 (CC)), where (x1 (CC), y1 (CC)) and (x2 (CC), y2 (CC)) represent the coordinates of the rightlower and left-upper bounder box corner. The punctuation marks are also deleted and the CC are categorized into two types (characters and words) depending on the space and width that lies between them. Pseudo-codes of connected components & punctuation removal are given in algorithms 2 and 3. Fig 4 presents the linked image components.

Algorithm 2: Detection of connected components

```
Input: Binary image
Output: Connected components
Begin:
label← 0
for i = 1 \rightarrow R do
    for j = 1 \rightarrow C do
       -\mathbf{i}\mathbf{\tilde{f}} I(i,j)is black pixel
          label←label+1
          Search(I,label,i,j)
        end if
    end for
 end for
 \\ coordinates of each labelled region
-For i =1 \rightarrow label do
  Find starting x and y coordinates of i^{th} labelled region
  Find width in pixels of i<sup>th</sup>labelled region
  Find height in pixels of i<sup>th</sup>labelled region
 end for
Search(I, label, i, j)
I(i,j) \leftarrow label
Find neighbours of I(i,j)
  For each neighbours (u, v) do
    [If neighbour (u, v) is black pixel
       Search(I,label, u ,v )
    end if
   end for
 end Search
end
```







Fig 4: Bounding box of connected components

### 4.3 Feature Extraction

An input document generally has a lot of information that is required for straight classification, but some of the information is not needed for the determination of the suitable class and as such, the pattern recognition process depends on original data extraction and measurement that is needed for classification. These objects are referred to features and features development presents a prime part of every domain of pattern recognition implementation. In this study, the author examines the top effective features for human personality analysis and samples classification in to various groups; for instance, solidity in handwriting refers to the high spiritual vitality of the writer, big letters refers to the outgoing and extroversion characteristics of the writer, and small letters refer to the reclusiveness and introverted character of the individual [45]. These features are used to categorize the handwriting documents into groups with a great level of similarity. The prime aim in this paper is to obtain the significant handwriting features for each passage, word, line and even characters, after which a group values is taken from each feature. The top six (6) features are detailed in the following paragraphs.

- 1. **Direction of Writing (f1)** this is gauged with the help of writing line baseline, indicating if the author's handwriting is rising lines, alignment lines, or dropping lines. Alignment line writing indicates that the writer is determined, ambitious, optimistic, and hopeful. The handwriting slant is also considered to be useful for writer style identification from other individuals as handwriting is akin to body language.
- 2. **Solidity (f2)** this gauge the handwriting intensity, reflecting each writer's style through the energy levels. A heavy pressure could indicate the quick reaction to criticism by the writer (or even perceived criticism). The intensity value ranges from 0 to 1, with those nearer to 0 being a potential irregular objective, and those nearer to 1 or 1, being a likely solid element.
- 3. Words Spacing (f3) this assists in writer's identification by indicating the class of the writer in that some need space more than others and a review of the samples of same writer and comparison with other writers showed that the distance between words can be useful in handwriting classification in light of the broad space, tight space, variation in space, among others.
- 4. **Gap between lines (f4)** this is gauged for the clarification of a robust attribute, recognizing the handwriting style of the individual as a plain space, very broad space, tight space or mixed lines, indicating good aim judgment.
- 5. Width of writing (f5) this constitutes the top features for classifying the handwriting style of a writer and is measured by black pixels tracing. Broad width is characterized by wider letters than they are tall, and narrow width by narrow letters than they are tall.
- 6. **Size of writing (f6)** this is an indication of the handwriting style of the person, where handwriting human personality analysis experts

find it useful to show the activity level, selfconfidence and self-importance of the individual. These may reflect the writer's aspiration, idealism, colonelcy, venture, bravery, generosity, glory, and need for space. Quite a big sized handwriting may indicate the exhibitionist nature of the writer and his craving for attention.

The details of the main features extraction is followed by the feature combination in order to determine the best vectors that symbolize every document.

# 4.4 Feature Grouping

Researches on writer classification mainly concentrated on the development or derivation of new features, while the author of the present study investigates the entire possibilities of features grouping to drop irrelevant features and to highlight the top ones. The features are indicated as f1, f2, f3, f4, f5, and f6. The procedure of feature grouping entails the creation of new features through the existing ones, describing input data to an optimum. The researcher examines and adopts seven features for human personality analysis.

This study focuses on the examination of the features behavior to achieve the effectiveness of combination, with all the potential features grouping represented as binary numbering system. In cases where the feature is represented as zero, the feature is excluded from the grouping and if it is represented as one that it is included, and as such, the possibilities initiate from the number one. The potential combination of (6) features are (26), equating to 64 groups of features.

# 4.5 Grouping

After the data specification of each input sample by a group of features, the samples are examined to determine the class or measurements to which they fall under. In the context of writer identification, the clusters are related to specific writers (every individual is modeled as the creator of specific handwriting patterns of a unique style. Such patterns examples are obtained from text images from the writing samples database and attributed to a specific writer. An example of the process of clustering is presented in fig 5.



#### Fig 5: Simple Example of Clustering Behaviour

The set of features in a cluster is used to provide a definition of the samples and this is known as the feature space, while the list of a single sample is known as feature vector. Every feature vector reflects a single point in the feature distance, and between the two vectors, there is a distance that indicates the samples similarity. The measurement of this distance is utilized to compare the component, using a biometric system. A set of features generate the feature vectors that are similar to the samples of a single person and different samples from other people. In an ideal scenario, the samples should be a single cluster for a specific person, and this should not overlap with other writers, which would lead to writer identification of an unknown document through the biometric system identification process that does the following;

- Calculates the text sample values for each feature
- Evaluates the group (cluster) it belongs to by determining the point it is closest to.

This study performed the generation process through various runs of different levels, with the first being features grouping (64 possibilities), and the second being the different distance measures used to calculate the distances among handwritten samples. These techniques of measurement include Correlation, Euclidean, Cityblock, Hamming, Cosine, Mahalanobis, Jaccard, Spearman, Chebychev and Minkowski. The third step involves the use of various clustering algorithms methods which are, average, ward, centroid, complete, median, single and weighted. In every process of individual clustering, the author used ten distance measures to generate various ensembles, and each generation mechanisms used the thresholds from 10 -100 increased by ten to generate partitions with various groups number. They are applied to determine and choose the top appropriate distance measurement and grouping techniques for the partitioning of the handwriting database samples effectively, with appropriate number of groups. The grouping process is akin to the persistent human information processing, with one of the aims behind using clustering algorithms being the development of automated tools for categorization or taxonomies data [46] [47]. The distance measures with their equations used in this study are listed in table 1.

Table 1: Distance measures with their expressions

Euclidean	$D_{Euc}(i) = \sqrt{\sum_{j=1}^{k} (V_j - M_{ij})^2}$
Hamming	$D_{Ham}(i) = \frac{Number of(V_j \neq M_{ij})}{k}$
Mahalanobis	$D_{Mah}(i) = (V - M_i)C^{-1}(V - M_i)'$
Cosine	$D_{\cos}(i) = \frac{\sum_{j=1}^{k} V_j M_{ij}}{\sqrt{\sum_{j=1}^{k} V_j^2} \sqrt{\sum_{j=1}^{k} M_{ij}^2}}$
Correlation	$D_{corr}(i) = \frac{(V - \overline{V})(M_i - \overline{M}_i)'}{\sqrt{(V - \overline{V})(V - \overline{V})'}\sqrt{(M_i - \overline{M}_i)(M_i - \overline{M}_i)'}}$
Cityblock	$D_{CB}(i) = \sum_{j=1}^{k} \left  V_j - M_{ij} \right $
Chebychev	$D_{Cheb}(i) = \max_{j} \left  V_{j} - M_{ij} \right $
Spearman	$\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2 - 1)}$
Jaccard	$d^{JAS}(i,j) = \frac{J11}{J01 + J10 + J11}$
Minkowski	$d^{MKD}(i,j) = \sqrt[\lambda]{\sum_{k=0}^{n-1} \left  y_{i,k} - y_{j,k} \right ^{\lambda}}$

The cluster framework for different phases is shown in fig 6. This determines the proper distance measures and features combinations in each individual cluster technique, possessing various numbers of classes.



Fig 6: Distance Measures for Different Number of Clusters

#### 5. FINDINGS AND DISCUSSION

This study aimed to study the effect of each single feature on the grouping methods and distance measure behaviors. Accordingly, the author focused on examining each feature individually and testing them for different features combination to determine

#### Academic Journal of Nawroz University (AJNU), Vol.10, No.1, 2021

the performance of features on the grouping systems and distance gauges. Different tests were carried out to evaluate the proposed technique performance and define the impact of various parameters for each characteristic and their grouping sets, using distance gauges and grouping algorithms. The work evaluation was conducted to determine appropriate distance measurements and grouping methods and the results were reached through the use of c=7 individual clustering (average linkage, ward linkage, centroid linkage, complete linkage, median linkage, single linkage and weighted linkage) for every distance measures and different groups. The study used different numbers of clusters to determine the performance of each cluster and the experiments are conducted using the KRDOH dataset of Kurdish handwritten text.

Based on the results of the experiments, enhancements were achieved by combining features from various feature groups (refer to Table 2). Specifically, features f2 and f6 were the highest performance over the group numbers increase in various distance measures and linkage techniques. This implies that features f2 and f6 have a significant influence on the performance of clustering systems and distance measures. The grouping of different features reflects significant performance enhancements in majority of the grouped sets.

Table 2: Features combination performance

Features	No. of	Grouping	Distance	F-
Combination	Groups	Method	Measures	Measure
f2, f6	10	average	chebychev	79.93
	10	centroid	cityblock	78.13
f2, f6, f3	10	single	cityblock	72.05
f2, f6, f1	10	single	cityblock	72.05
	10	single	euclidean	72.04
f6, f5, f3	10	single	euclidean	72.04
f6, f1, f2	10	single	cityblock	72.04
f6, f1, f5	10	single	cityblock	72.04
f6, f3, f1	10	single	chebychev	72.04
f6, f3, f1, f5	10	single	chebychev	72.04
f4, f2, f3, f5	10	single	cosine	72.04
f4, f6, f3	10	median	jaccard	72.04
f6, f3, f1	10	median	hamming	72.04
f6, f3, f1, f5	10	median	hamming	72.04
f2, f6, f5	10	median	hamming	72.04
f2, f6, f3, f1	10	single	hamming	72.04
f2, f6, f3, f1, f5	10	single	hamming	72.04
f1, f5	30	average	hamming	72.03
f1, f5	40	average	hamming	72.01
f1, f5	50	centroid	hamming	72.00
f4, f2, f6, f1, f5	10	single	correlation	72.00
f4, f2, f6, f3, f1, f5	10	single	correlation	72.00
f1, f5	60	average	hamming	71.96
f1, f5	70	centroid	hamming	71.95
f1, f5	80	average	hamming	71.90
f1, f5	90	single	hamming	71.89
f1, f5	100	average	spearman	71.84
f4, f2, f6, f3	100	single	mahalanobis	70.71
f6, f3	100	single	euclidean	70.71
f6, f3	100	single	minkowski	70.71
f4, f2, f6, f3, f1, f5	100	single	mahalanobis	70.70

Fig 7 illustrates the overall performance of clustering methods in multiple run. Average, Centroid, median and single cluster methods exhibit stable performance with increasing the number of clusters. However, the other clustering methods performance is appeared to reduce with the increase of clusters number.



Fig 7: Overall Performance of Grouping Methods in Multiple Run.

### 6. CONCLUSIONS

This work proposed a new clustering approach of Kurdish handwritten documents adopting the top significant features and using them in the same methods used in the human characteristics investigate for handwriting. Each feature's impact was studied using clustering methods and distance gauges behaviors, examining every individual feature and testing the combination of features sets to determine the performance of features on the grouping systems and distance gauges, in light of different groups numbers. Based on the findings, the graph-based multiple runs for grouping methods and distance measurement had an excellent performance, with values of f-measure similar in different numbers of groups for the majority of features. Also, the top and stable performance of groups methods was generated by single linkage over all features and the performance of majority of distance gauges (i.e., Mahalanobis, Euclidean, Hamming, Jaccard, Minkowsi, Spearman, among others) were consistent and depended on the clustering methods, features and the clusters number, with the Spearman's performance performing better compared to other distance measures. It provides an almost stable value with the increase of the clusters numbers. On the whole, the optimum result was reached when intensity, lines slope and combination of set of features were employed, indicating that feature significantly impacts the performance of grouping systems and distance gauges. The findings indicate that grouping performance effectiveness can be enhanced through the graph-based individual grouping, albeit such enhancement broadly relies on the ensemble mechanism and the features utilized.

## ACKNOWLEDGMENT

The author would like to thank Qaiwan International University (QIU) for supporting this

research. We also thank the anonymous reviewers whose comments have improved this paper. **REFERENCES** 

- S. Srihari, C. Huang, H. Srinivasan, and V. Shah, "Biometric and forensic aspects of digital document processing," in *Digital Document Processing*, 2007, pp. 379–405.
- B. Zhang and S. N. Srihari, "Analysis of handwriting individuality using word features," in Seventh International Conference on Document Analysis and Recognition, Proceedings., vol. 1, pp. 1142–1146, 2003.
- 3. S. Srihari, S. Cha, H. Arora, and S. Lee, "Individuality of handwriting," Journal of forensic science, 47(4), 1-17, 2002.
- 4. S. Srihari, S.-H. Cha, and S. Lee, "Establishing handwriting individuality using pattern recognition techniques," In Proceedings of Sixth International Conference on Document Analysis and Recognition pp. 1195–1204, 2001.
- S. Srihari, H. Arora, S. N., S.-H. Cha, and S. Lee, "Individuality of handwriting: a validation study," in Proceedings of Sixth International Conference on Document Analysis and Recognition, 2001, pp. 106–109.
- Y. Zhu, T. Tan, and Y. Wang, "Biometric personal identification based on handwriting," In Proceedings 15th International Conference on Pattern Recognition. ICPR-2000 (Vol. 2, pp. 797-800). IEEE, 2000.
- BBC, "Graphology," Oxford, 2009. [Online]. Available: http://www.bbc.co.uk/oxford/content/articles/ 2007/05/23/handwriting\_sybil.shtml.
- Graphology Handwriting Analysis, [Online]. Available: http://www.businessballs.com/graphologyhand

writinganalysis.htm.

- C.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: investigation of normalization and feature extraction techniques," Pattern Recognit., vol. 37, no. 2, pp. 265–279, Feb. 2004.
- C.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: benchmarking of state-of-the-art techniques," *Pattern Recognit.*, vol. 36, no. 10, pp. 2271–2285, Oct. 2003.
- Xu, D. Y., Shang, Z. W., Tang, Y. Y., & Fang, B. (2008, August). Handwriting-based writer identification with complex wavelet transform. In 2008 International Conference on Wavelet Analysis and Pattern Recognition (Vol. 2, pp. 597-601). IEEE.
- 12. A. Schlapbach and H. Bunke, "Using HMM Based Recognizers for Writer Identification and

Verification," In Ninth International Workshop on Frontiers in Handwriting Recognition, pp. 167– 172, 2004.

- 13. Bensefia, T. Paquet, and L. Heutte, "A writer identification and verification system," Pattern Recognit. Lett., vol. 26, no. 13, pp. 2080–2092, Oct. 2005.
- 14. K. Yu, Y. Wang, and T. Tan, Biometric Authentication, vol. 3072. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.
- 15. Z. Y. HE and Y. Y. Tang, "Chinese handwritingbased 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, 2004.
- Shen, X. Ruan, and T. Mao, "Writer identification using Gabor wavelet," in Proceedings of the 4th World Congress on Intelligent Control and Automation (Cat. No.02EX527), vol. 3, pp. 2061– 2064, 2002.
- U. Marti, R. Messerli, and H. Bunke, "Writer identification using text line based features," in Proceedings of Sixth International Conference on Document Analysis and Recognition, pp. 101–105, IEEE, 2001.
- 18. A. Koerich, "Large vocabulary off-line handwritten word recognition," École de technologie supérieure, Aug. 19, 2002.
- 19. A. Koerich, R. Sabourin, and C. Suen, "Large vocabulary off-line handwriting recognition: A survey," Pattern Anal. Appl., vol. 6, no. 2, pp. 97–121, Jun. 2003.
- 20. R. Cajote and R. Guevara, "Global word shape processing using polar-radii graphs for offline handwriting recognition," in 2004 IEEE Region 10 Conference TENCON, vol. A, pp. 315–318, 2004.
- 21. M. Morita, "Automatic Recognition of Handwritten Dates on Brazilian Bank Cheques," 2004.
- 22. S. Madhvanath and V. Govindaraju, "The role of holistic paradigms in handwritten word recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 2, pp. 149–164, 2001.
- 23. S. Fatimah, C. Djamal, R. Ilyas, and F. Renaldi, "Personality Features Identification from Handwriting Using Convolutional Neural Networks," In 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), pp. 119-124, IEEE, 2019.
- 24. Fallah, and H. Khotanlou, "Identify human personality parameters based on handwriting using neural network," In 2016 Artificial Intelligence and Robotics, (IRANOPEN), pp. 120-126, IEEE, 2016.

- 25. S. Hemlata, and K. Singh, "Personality Detection using Handwriting Analysis," In The Seventh International Conference on Advances in Computing, Electronics and Communication (ACEC), pp. 85-89, 2018.
- 26. N. Lakshmi, and A. Keerthana, "Handwriting analysis based human personality prediction using sugeno fuzzy model," Int. J. Sci. Eng. Res., 8(5), 105-110, 2017.
- 27. Z. He, X. You, and Y. Tang, "Writer identification using global wavelet-based features," Neurocomputing, vol. 71, no. 10–12, pp. 1832– 1841, Jun. 2008.
- 28. G. Leedham and S. Chachra, "Writer identification using innovative binarised features of handwritten numerals," in Seventh International Conference on Document Analysis and Recognition, Proceedings., vol. 1, pp. 413–416, 2003.
- 29. E. Zois and V. Anastassopoulos, "Morphological waveform coding for writer identification," Pattern Recognit., vol. 33, no. 3, pp. 385–398, Mar. 2000.
- Doermann, "Indexing of handwritten historical documents-recent progress," In Symposium on document image understanding technology, pp. 77-85, 2003.
- T. Rath and R. Manmatha, "Word image matching using dynamic time warping," In 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings., vol. 2, pp. II-521-II-527, 2003.
- 32. S. Kane, A. Lehman, and E. Partridge, "Indexing George Washington's handwritten manuscripts," Center for Intelligent Information Retrieval, Computer Science Department, University of Massachusetts, Amherst, MA 1003, 2001.
- 33. M. Rath, S. Kane, A. Lehman, E. Partridge, and R. Manmatha, "Indexing for a digital library of George Washington's manuscripts: a study of word matching techniques," CIIR Technical Report, 2002.
- 34. Tomai and V. Govindaraju, "Transcript mapping for historic handwritten document images," In Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition, pp. 413– 418, 2002.
- 35. M. Rath, V. Lavrenko, and R. Manmatha, "Retrieving Historical Manuscripts Using Shape," 2003.
- 36. H. Al-Rashaideh, "Preprocessing phase for Arabic word handwritten recognition," Information Process (Russian) 6, no. 1 (2006).

- 37. Trier, K. Jain, and T. Taxt, "Feature extraction methods for character recognition-a survey," Pattern recognition, 29(4), 641-662, 1996.
- 38. M. Khorsheed and W. Clocksin, "Structural Features of Cursive Arabic Script," Proceedings Br. Mach. Vis. Conf. 1999, pp. 1–10, 1999.
- 39. H. Alazim, "A hybrid fuzzy-neural approach to the recognition of arabic script," In Proceedings of the 5th International Conference and Exhibition on Multi-Lingual Computing, p. 2, 1914.
- H. Cha, and N. Srihari, "A priori algorithm for sub-category classification analysis of handwriting," In Proceedings of Sixth International Conference on Document Analysis and Recognition, pp. 1022-1025, IEEE, 2001.
- 41. K. Powalka, N. Sherkat, and J. Whitrow, "Recognizer characterisation for combining handwriting recognition results at word level," In Proceedings of 3rd International Conference on Document Analysis and Recognition, Vol. 1, pp. 68-73, IEEE, 1995.
- A. Mostafa, M. Al-Qurishi, and I. Mathkour, "Towards Personality Classification Through Arabic Handwriting Analysis," In The International Research & Innovation Forum, Springer, Cham, pp. 557-565, 2019.
- 43. M. Gavrilescu, N. Vizireanu, "Predicting the big five personality traits from handwriting," (EURASIP), Journal on Image and Video Processing, no. 1, p. 57, 2018.
- 44. T. Latif, A. Abdullah Ahmed, and O. Ismael Al-Sanjary, "KRDOH: Kurdish Offline Handwritten Text Database," In 2019 IEEE 7th Conference on Systems, Process and Control (ICSPC), pp. 86–89, 2019.
- 45. A. Furnham and B. Gunter, "Graphology and personality: Another failure to validate graphological analysis," Pers. Individ. Dif., vol. 8, no. 3, pp. 433–435, Jan. 1987.
- 46. N. Jardine and R. Sibson, "Mathematical taxonomy," London etc. John Wiley, 1971, [Online]. Available: http://kbd.kew.org/kbd/detailedresult.do?id=3 8228.
- 47. P. Sneath and R. Sokal, "Numerical taxonomy. The principles and practice of numerical classification.," 1973,
- 48. P.R. Christopher D. Manning, H. Schutze, Introduction to Information Retrieval, 2008.
- G.M. Weiss, F.J. Provost, Learning When Training Data are Costly: The Effect of Class Distribution on Tree Induction, J. Artif. Intell. Res. 19 (2003) 315– 354.