New Online/Offline text-dependent Arabic Handwriting dataset for Writer Authentication and Identification

1st Mohammad Z. Al-Shamaileh
Dept. of Information Technology
Mutah University
Karak, Jordan

2nd Ahmad B. Hassanat
Dept. of Information Technology
Mutah University
Karak, Jordan

3rd Ahmad S. Tarawneh
Dept. of Algorithms
and Their Applications
Eötvös Loránd University
Budapest, Hungary

4thM. Sohel Rahman
Dept. of CSE , BUET,
ECE Building, West Palasi,
Dhaka 1205, Bangladesh

5th Ceyhun Celik
Department of Computer Engineering
Gazi University
Ankara, Turkey

6th Moohanad Jawthari
Dept. of Algorithms
and Their Applications
Eötvös Loránd University
Budapest, Hungary

***Abstract*—Word-based writer identification and authentication have been under investigation for years. In this paper, we present a new Arabic online/offline handwriting dataset for writer authentication and identification. The created dataset includes two parts: AHWDB1 and AHWDB2 which are made freely available for the research community. Each part of the dataset consists of 2000 (10 trials X 200 writers) samples captured using HUAWEI MediaPad M3. Dynamic information such as pressure, timestamp and the coordinates(X, Y) have been collected and involved for both parts of the dataset. In addition, age, gender and education level have been added to the dataset for future investigation. Several experiments are conducted on the dataset. The preliminary identification results are obtained using K-Nearest Neighbor (KNN) with Dynamic time warping (DTW) and support vector machines (SVM). We recorded 81.35% identification rate as the best result using SVM classifier on AHWDB2.**

***Index Terms*—Behavioral biometric, KNN, SVM, Machine learning, DTW, AI.**

I. INTRODUCTION

The increasing use of smart phones and tablets raises the need for simple and secure user authentication and identification systems [1], as smartphones have limited capabilities of processing, storage and security [2]. Such limitations can be overcome using biometrics as an initiative alternative especially for behavioral events such as handwriting. Touch screen is the only required hardware for handwriting identification and authentication systems. It is mainly available in most smartphones and tablets, unlike other biometrics such as fingerprint, iris, retina, etc., which require special hardware to extract the relevant information [3]. While it is true that many of the recent devices do provide fingerprint facilities, it is still not common. Handwriting as a behavioral biometric varies from one writer to another [4]. The same writer cannot write a text exactly the same each time, however, they can write it with some approximation, which can be measured with some similarity measure [5]. Therefore, handwriting can be used as a behavior biometric. The language being used is also an important factor for the authentication and identification system due to the differences among languages in various characteristics, including, but not limited to, characters, cursive or printing, rules, the direction of writing, etc. [6].

Based on the input method, there are two approaches of recognition, namely, online and offline [7]. Online handwriting recognition is concerned about the direct acquisition of the dynamic information for the written script as it is being written on the device. offline recognition approaches, on the other hand, acquire static information of an image of the written script, which is usually scanned from papers. However, the recognition accuracy of an offline approach is usually lower than that of an online one due to the absence of the rich dynamic information [8] in the former approach.

Writer identification approaches are classified into two categories based on the text content [9]. In the text-dependent approach, the writer is required to write an exact text many times. The text-independent approach, however, does not force the writer to write a specific text, usually a few lines of words.

In general, recognizing a writer based on his/her handwriting, in any language, is a behavioral biometric [10]. It depends on the fact that the style of writing is an educated and early age personal habit and thus, technically, the way of identifying a person from his/her handwriting resembles other biometric-based recognition. In its core, a recognition method needs to extract meaningful features that can effectively represent handwriting so as to distinguish among the handwriting of different writers. Furthermore, a proper classification scheme is also required. While such methods are available for other languages such as English, while for Arabic handwriting, there are multiple issues that need special attention. For example, in the case of Arabic handwriting, we need to consider the differences in structure and characteristics when dealing with identification challenges [11].

There are only a few studies in the literature on online text-dependent writer identification as opposed to many works on text-independent schemes [9], [12]–[15]. Also, to the best of our knowledge, there is no text-dependent online handwriting dataset exists in the literature. In fact, while reviewing the literature, we did not find any online text-dependent writer identification scheme that worked on the Arabic language.

The main contribution of this paper is as follows. Firstly, we have created two Arabic handwriting datasets, namely, Arabic handwriting database (AHWDB1) and AHWDB2, which can be used to evaluate methods of identifying writers by their Arabic handwriting from one or two words only. AHWDB1 and AHWDB2 are made freely available through Mutah University website (https://www.mutah.edu.jo/biometrix/) for research purposes. Secondly, we propose a method for online Arabic text-dependent writer identification and evaluate its effectiveness and efficacy using the above two datasets.

II. Literature Review

Due to its importance in some fields such as biometric authentication in information and network security, document analysis systems, etc. [16], A lot of efforts have been made in the recent years on identifying a person from his/her handwriting. Numerous researches have been conducted in the field of writer identification for a variety of different languages [17]. Researchers adapted many approaches that target a specific category of identification such as online and offline, or based on the content of the text, such as text-dependent and text-independent [18]. Most researches extracted the features from characters, words, lines, or pages. Other researches extracted the features from smaller parts such as diacritics, small strokes, or parts of characters. Different handwriting languages have been tested by these approaches [17].

An offline text-independent writer identification system using bagged discrete cosine transform (BDCT) descriptors was proposed in [18]. In this approach, multiple predictor models were first generated using universal codebooks and majority voting was then used on these predictor models to obtain the final decision. The researchers claimed that the effective usage of the discrete cosine transform features was the reason for the robustness of their system. Four datasets (IFN/ENIT [19], AHTID/MW [20], CVL [21] and IAM [22]) were used to evaluate the performance of their proposed system. Results were comparable with existing ones in the field. Also, extensive experiments on blurry and noisy images of two different datasets (AHTID/MW and 100 randomly selected writers from the IAM) were conducted and improved performance of their proposed system was reported with respect to the existing systems.

A novel approach was proposed in [23] for writer identification using Arabic offline text-independent handwritten images. Local Binary Pattern (LBP) was used as a texture descriptor in this method. Also, K-Nearest Neighbor (KNN) with Hamming distance was applied on a part of IFN/ENIT dataset (50-word images from 130 writers) [19] and the proposed method achieved an accuracy of 87%.

A text-independent online Arabic writer identification method was proposed by [24]. The study was based on a variety of dynamic and statistical features extracted from stroke speed, the distance between strokes, point-based features, word-based features, and horizontal histogram were used at different levels in the word. Dynamic Time Warping (DTW) and Support Vector Machine (SVM) were utilized for classification. ADAB DB (33000 words written by 166 writers) [25] was used to evaluate the approach; however, the results were not satisfactory with only 45.67% identification rate for 19 writers and 30.05% for 41 writers.

An automatic text-independent online Arabic writer identification system was proposed in [26]. After extracting a set of features (Dynamic, and Geometric) by the Beta-Elliptic Model (BEM), a feed-forward neural network was applied as the classifier and the proposed system was evaluated on a subset of the ADAB dataset [27]. The results were excellent with 91.22% identification rate for the top 1 and 100% for the top 5.

In order to enhance the performance of writer identification systems [12] proposed a combined method of both online and offline approaches based on multi-fractal features. The authors used several methods to extract the multi-fractal features as follows. For offline features, they used Diffusion Limited Aggregates (DLA), Box-Counting, Average and Box-Counting Density (ABCD). On the other hand, for online features, they utilized DLA-ON, ABCD-ON, and compass. Finally, KNN was used for classification. The system was evaluated on the writings of 110 writers from ADAB dataset [27]. For the offline case, accuracy rates of 70.3% and 85.9% were achieved for Top-1 and Top-10, respectively. For the online case, 73.1% and 86.5% accuracy rates were attained; a combination of both approaches resulted in 83.8% and 91.3% accuracy rates.

III. The Proposed Method

The proposed method consists of three major steps, namely, handwriting acquisition, preprocessing and machine learning. Firstly, a sequence of coordinates (X, Y), which represents the pixels of the handwritten word(s) on the tablet is generated synchronously during the user enrolment. Secondly, these coordinates are preprocessed to make the samples invariant to rotation, scale and translation. The new preprocessed coordinates are either mapped (to standardize the different lengths of the samples) to a standard matrix size, or left as they are. Thirdly, a machine learning technique is applied for matching. A block diagram of the proposed method is illustrated in Figure 1.

A. Handwriting acquisition

In this step, the handwritten text is captured by tablets. The user writes the requested word(s) on the tablet's touch screen. The coordinates of the pixels of the written word(s) on the screen are registered directly (online) and will be used later.

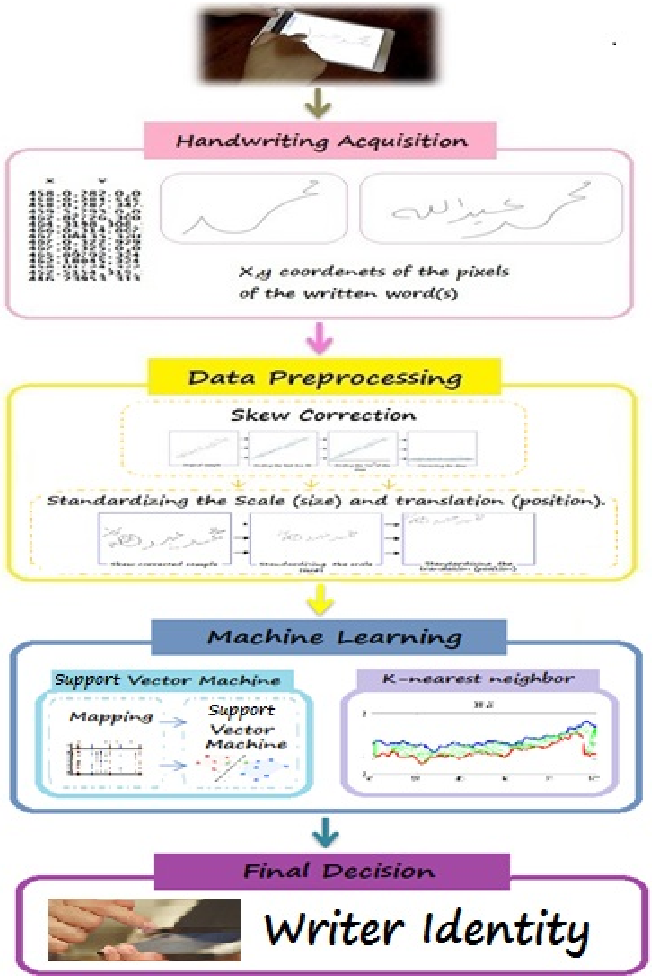

Fig. 1. A block diagram of the proposed method

Fig. 2. Eight samples of different writers from AHWDB1 (bottom row) and AHWDB2 (top row) with variations on rotation, scale (size), and translation (position).

Note that this is an online approach as information of the written text is immediately registered during the writing process. In particular, various dynamic information like timestamp, X coordinate, Y coordinate, and pressure are recorded. For each touch, these data are saved in a text file.

B. Preprocessing

Preprocessing is an essential step for the purpose of improving the accuracy of the identification process due to the variations in scale, position, and rotation of the handwritten samples among the different writers. Some degrees of variation also exist within the same writer's samples (see Figure 2).

Although line regression is one of the well-known prediction methods, it can also be used for skew correction due to its ability to find the best line fit between a set of points. Here, we use the simple line regression to detect the skewed written word(s) by finding the best line that crosses the pixels thereof, and subsequently correct those by rotating the word(s) according to the angle of the slope as illustrated in Figure 3.

Generally, in most biometric systems, the difference in scale (size), and translation (position) among samples can lead to an inaccurate identification. Therefore, we have made an effort to standardize the scale (size), and translation (position) of the samples through a normalization process (i.e., transforming the coordinates to be within a specific range). This also prepares the coordinates for the next stage, i.e., mapping. There are various normalization techniques to choose from, including, but not limited to, min-max, z-score, Tanh-estimator, etc. [28], [29]. We have opted for the min-max normalization for the sake of simplicity.

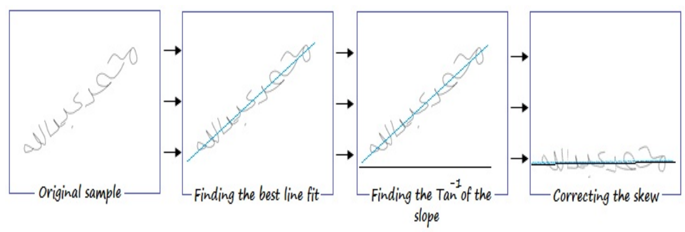

Fig. 3. Skew detection and correction using simple line regression.

C. machine learning

We have used two methods for the matching stage: a) the KNN with Dynamic Time Warping (DTW) algorithm, which is used on the coordinates that represent each word, and b) the Support Vector Machine (SVM) after mapping the coordinates into an arbitrarily fixed size matrix, which is used to make the size of the feature vectors uniform, and to enhance the process as well.

1) K-Nearest Neighbor (KNN) with DTW: A meta-analysis conducted by [30] suggests that KNN with k=1 and DTW is one of the best classifications for time series data. Typically, in writer recognition systems, the feature vectors are of different lengths, and therefore cannot be compared by the common similarity measures such as Euclidean, or Manhattan distances [31]. Thus DTW emerges as one of the choices that can be used in our setting.

2) Support Vector Machine with Mapping: In this method, the normalized coordinates (in range 0-1) are mapped into a two-dimensional binary matrix of size M x N. These dimensions were selected based on preliminary experiments. Due to the variations in the feature vector size, we need a faster matching method than the DTW, which has quadratic time complexity. Thus, due to its fixed vector size, mapping is a good alternative here requiring O(K) time only in the matching stage, where k is the constant size of the mapped matrix (M X N).

This binary matrix is supposed to capture the representation of the written words, where 1s represent the locations of the points forming the word, and 0s represent empty spaces. To conduct the mapping, the normalized X-coordinates are multiplied with M, and rounded up to fit a location within the matrix. The same is undertaken for the Y-coordinates with N. M and N are empirically assigned (see results section). Based on the new location, ones are assigned for each point, thus a fixed size matrix of zeros and ones is obtained. This matrix forms the feature vector for each written word, which will be used for the matching process.

After preparing the new feature vector (mapped matrix), SVM is applied as the classifier. Notably, we have done preliminary experiments with several other classifier algorithms and found SVM to be the best performer.

IV. DATASET PREPARATION

Table 1 presents the characteristics of the datasets that are already available in the literature.

Recall that our goal is to identify writers by their Arabic handwriting from one or two words only. The lack of an appropriate dataset to properly evaluate our proposed method prompted us to prepare a new dataset, which will be useful for other researchers as well. In particular, we have created two Arabic handwriting datasets, namely, AHWDB1 and AHWDB2. To prepare these datasets, every subject sat on a chair with a tablet in front of him/her on a table, and was asked to write with his/her finger horizontally from right to left. In AHWDB1 dataset, each subject wrote the word, Mohammad (محمد), repeated 10 times. On the other hand in AHWDB2 dataset, two words, ”Mohammad Abdallah” (محمد عبدالله) were written by the subject 10 times on the same tablet. The reason for selecting these words is that the word Mohammad is the most common name in Arabic. The 200 writers were requested to take a short break of approximately 10 seconds before each trail, so each writer wrote 20 trails in one session.

Thus we get two datasets, where each one consists of 2000 samples (10 trails x 200 writers) representing the handwriting of the same 200 writers (males and females). The ages of the writers were in the range of 9 to 63 years, and their educational level varied from basic education (school) to Doctoral degree. All writers wrote with their right hand except for 5 of them who were left-handed. The writers age, gender, education level and which hand the writer used to write with are added as a part of the dataset.

HUAWEI MediaPad M3 was used for the acquisition of the handwriting. Specifications of the device are shown in Table 2. A special Android application was developed and installed for the acquisition process. The application was user-friendly and had the capability of clearing the screen and rewriting when required. Every trail's information (X-coordinate, Y-coordinate, timestamp, pressure) was stored in a text file (as shown in Figure (4)) with a name defined through a concatenation of the user serial number and the corresponding trail number (e.g., User 22 and trial 14: (22-14)). Also, an image (JPEG format) of the handwritten word on the screen is saved. Notably, in this research, we only use the X-coordinate and Y-coordinate information for training and classification purposes, leaving other information (pressure and time) for future work.

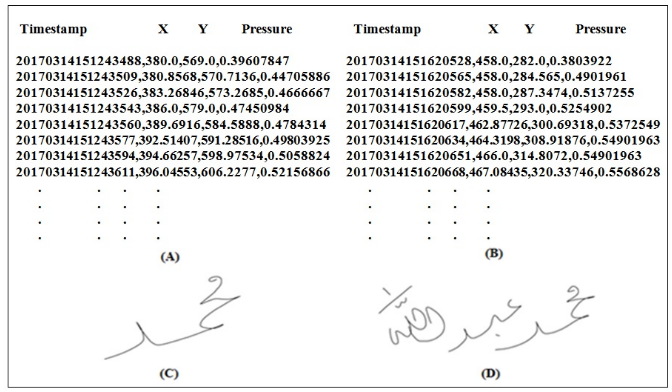

Fig. 4. (A) timestamps, X coordinates, Y coordinates and pressure information of trail 6 from subject 195 from the AHWDB1 dataset. (B) Same information of trail 14 from subject 195 from the AHWDB2 dataset. (C) An image of (A). (D) an image of (B).

V. RESULTS AND DISCUSSIONS

A. Results based on KNN with DTW

We have evaluated KNN with DTW using our datasets. K is set to equal 1 and Manhattan distance is used to perform DTW. Leave-One-Out cross validation approach is used for the evaluation.

A separate experiment was conducted for each part of the dataset with the same method. The X-coordinates and Y-coordinates were the feature vectors that represented each sample. As seen in Table 3, the identification rate achieved using AHWDB2 is higher than that achieved using AHWDB1. This may be attributed to the fact that ”Mohammad Abdallah” is longer than ”Mohammad” in its length which supports the assumption that the longer the word the higher the accuracy ratio will be. Accuracy here refers to the number of correctly classified instances divided by the total number of instances.

B. Results based on SVM with Mapping

Mapping standardizes the size of the feature vectors (the X-coordinates and Y-coordinates) which offers the opportunity to utilize other measurement metrics. The choice of the size of the new feature vectors (matrix of zeros and ones) is of critical importance. To identify the best-performing size of the map, several experiments were performed on both datasets with the new mapped feature vectors using SVM.

In parallel to applying SVM, we applied other classifiers using WEKA workbench with their default parameter values

TABLE I
THE CHARACTERISTICS OF THE DATASETS THAT MENTIONED IN THE LITERATURE.

Dataset	Writers	Content	Language
IFN/ENIT [19]	411	26400 names containing more than 210000 characters	Arabic
IAM [22]	400	82227 words	English
CVL [21]	311	101069 words	English German
AHTID/MW [20]	53	22896 words	Arabic
ADAB DB [25]	More than 130	15158 words	Arabic

TABLE II
HUAWEI MEDIAPAD M3 SPECIFICATIONS.

launch	Announced	2016, September
	Status	Available Released 2016, October
body	Dimensions	215.5 x 124.2 x 7.3 mm
	Weight	-
	SIM	Nano-SIM
Display	Type	IPS LCD capacitive touchscreen
	Size	8.4 inches, 204.6 cm2
	Resolution	1600 x 2560 pixels, 16:10 ratio
	Multitouch	Yes
		-EMUI 4.1
Platform	OS	Android 6.0 (Marshmallow)
	Chipset	Hisilicon Kirin 950
	CPU	Octa-core (4x2.3 GHz Cortex-A72 & 4x1.8 GHz cortex A53)
	GPU	Mali-T880 MP4
MEMORY	Card slot	microSD, up to 256 GB (dedicated slot)
	Internal	32/64 GB, 4GB RAM

TABLE III
IDENTIFICATION RESULT OF APPLYING KNN WITH DTW ON AHWDB1 AND AHWDB2.

dataset	Accuracy
AHWDB1	0.7375
AHWDB2	0.7535

[32], namely, random forest (10 folds), KNN (2000 folds), neural network (10 folds), and SVM (10 folds) on both AHWDB1 and AHWDB2. As illustrated in Table 4, SVM with 10 folds and its default parameters has achieved the highest identification rate 81.35% on AHWDB2, while we found that the DTW approach performed better than the mapping approach on the AHWDB1 dataset, this is because the raw feature vector of the AHWDB2 is larger than that of the AHWDB1, which allows variations in the DTW and at the same time fill more gaps in the map.

The best results were achieved when the size of the map was 12 x 10 with 72.35% for the AHWDB1 and 18 x

TABLE IV
ACCURACY RATES USING DIFFERENT CLASSIFIERS ON AHWDB1 AND AHWDB2.

Classifier	AHWDB1 Accuracy	AHWDB2 Accuracy
ANN	0.6980	0.7825
SVM	**0.7235**	**0.8135**
RF	0.6120	0.6735
KNN	0.6690	0.7490

AHWDB1				AHWDB2			
Size			Accuracy	Size			Accuracy
10	x	12	0.7125	13	x	13	0.8010
12	x	15	0.7125	15	x	15	0.8010
10	x	17	0.7130	14	x	22	0.8015
9	x	21	0.7135	15	x	19	0.8015
9	x	25	07135	18	x	19	0.8015
13	x	12	0.7135	12	x	20	0.8020
9	x	13	0.7140	16	x	15	0.8020
12	x	12	0.7140	18	x	13	0.8025
14	x	18	0.7140	12	x	19	0.8030
15	x	12	0.7040	13	x	16	0.8030
12	x	16	0.7155	18	x	14	0.8035
9	x	16	0.7160	13	x	17	0.8055
13	x	16	0.7160	16	x	13	0.8060
10	x	16	0.7165	12	x	18	0.8065
14	x	12	0.7165	13	x	19	0.8075
14	x	16	0.7180	14	x	18	0.8075
12	x	13	0.7195	13	x	22	0.8090
11	x	16	0.7210	13	x	20	0.8105
9	x	17	0.7220	13	x	18	0.8120
12	x	10	0.7235	18	x	15	0.8135

15 with 81.35% for the AHWDB2. Table (5) illustrates the identification rates using different sizes of feature vectors (best 20 results from testing all sizes from 5 x 5 to 40 x 40).

As can be seen from Table 5, it is interesting to note that the size of the map is proportional to the size of the raw feature vector, for the smaller (one-word) dataset (AHWDB1), the best Map size is in the range (11 x 11)2, while for the larger (two-word) dataset (AHWDB2), the best Map size is in the range (16 x 16)2.

VI. CONCLUSION

A new and simple method for online text-dependent writer identification has been proposed in this paper. At first, we do some necessary preprocessing, followed by KNN with DTW or the SVM with mapping (i.e., the X and Y coordinates were put into an N x M matrix).

Two new Arabic online/offline text-dependent handwriting datasets were created to address the problem of writer authentication and identification because no applicable dataset was available in the literature. We have made these datasets available online with the hope that more investigation in the field of Arabic handwriting will be undertaken.

The experiments show a better accuracy rate of identification, using SVM with Mapping than KNN with DTW on AHWDB2, and a close performance between the two classifiers on AHWDB1. Notably, SVM with Mapping works faster than KNN with DTW. However, there is still much scope for further improvement. It would appear that extracting new features or to use the unused features such as pressure or/and timestamp could potentially enhance the identification rate. This limitation will be addressed in our future work, in addition to using other methods for features extraction such as [33] and deep features [34]. Moreover, other KNN-based classifiers will be investigated such as [35]–[37].

Acknowledgment

The third author would like to thank Tempus Public Foundation for sponsoring his Ph.D. study, also, his work is under the project EFOP-3.6.3-VEKOP-16-2017-00001 (Talent Management in Autonomous Vehicle Control Technologies), and supported by the Hungarian Government and co-financed by the European Social Fund.

References

[1] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello, "Continuous user authentication on mobile devices: Recent progress and remaining challenges," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 49–61, 2016.

[2] S. Zonouz, A. Houmansadr, R. Berthier, N. Borisov, and W. Sanders, "Secloud: A cloud-based comprehensive and lightweight security solution for smartphones," *Computers & Security*, vol. 37, pp. 215–227, 2013.

[3] T. Sabhanayagam, V. P. Venkatesan, and K. Senthamaraikannan, "A comprehensive survey on various biometric systems," *International Journal of Applied Engineering Research*, vol. 13, no. 5, pp. 2276–2297, 2018.

[4] M. Topaloglu and S. Ekmekci, "Gender detection and identifying one's handwriting with handwriting analysis," *Expert Systems with Applications*, vol. 79, pp. 236–243, 2017.

[5] R. A. Huber and A. M. Headrick, *Handwriting identification: facts and fundamentals*. CRC press, 1999.

[6] A. Hassanat and G. A. Altarawneh, "Rule-and dictionary-based solution for variations in written arabic names in social networks, big data, accounting systems and large databases," *arXiv preprint arXiv:1502.05441*, 2015.

[7] A. Priya, S. Mishra, S. Raj, S. Mandal, and S. Datta, "Online and offline character recognition: A survey," in *2016 International Conference on Communication and Signal Processing (ICCSP)*. IEEE, 2016, pp. 0967–0970.

[8] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of biometrics*. Springer Science & Business Media, 2007.

[9] T. Dhieb, W. Ouarda, H. Boubaker, and A. M. Alimi, "Deep neural network for online writer identification using beta-elliptic model," in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 1863–1870.

[10] 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. IEEE, 2000, pp. 797–800.

[11] B. M. Al-Helali and S. A. Mahmoud, "Arabic online handwriting recognition (aohr): a survey," *ACM Computing Surveys (CSUR)*, vol. 50, no. 3, p. 33, 2017.

[12] A. Chaabouni, H. Boubaker, M. Kherallah, H. El Abed, and A. M. Alimi, "Static and dynamic features for writer identification based on multi-fractals." *Int. Arab J. Inf. Technol.*, vol. 11, no. 4, pp. 416–424, 2014.

[13] A. Namboodiri and S. Gupta, "Text independent writer identification from online handwriting," in *Tenth International Workshop on Frontiers in Handwriting Recognition*. Suvisoft, 2006.

[14] A. Chaabouni, H. Boubaker, M. Kherallah, A. M. Alimi, and H. El Abed, "Combining of off-line and on-line feature extraction approaches for writer identification," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 1299–1303.

[15] V. Venugopal and S. Sundaram, "An online writer identification system using regression-based feature normalization and codebook descriptors," *Expert Systems with Applications*, vol. 72, pp. 196–206, 2017.

[16] A. Raj and A. Chaudhary, "Personnel identification using handwriting, tested on indian writers," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 5, no. 6, pp. 232–235, 2017.

[17] A. Raj and A. Chaudhary, "A review of personal identification using handwriting," *Int J Eng Sci Comput*, 2016.

[18] F. A. Khan, M. A. Tahir, F. Khelifi, A. Bouridane, and R. Almotaeryi, "Robust off-line text independent writer identification using bagged discrete cosine transform features," *Expert Systems with Applications*, vol. 71, pp. 404–415, 2017.

[19] M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, H. Amiri *et al.*, "Ifn/enit-database of handwritten arabic words," in *Proc. of CIFED*, vol. 2. Citeseer, 2002, pp. 127–136.

[20] A. Mezghani, S. Kanoun, M. Khemakhem, and H. El Abed, "A database for arabic handwritten text image recognition and writer identification," in *2012 international conference on frontiers in handwriting recognition*. IEEE, 2012, pp. 399–402.

[21] F. Kleber, S. Fiel, M. Diem, and R. Sablatnig, "Cvl-database: An off-line database for writer retrieval, writer identification and word spotting," in *2013 12th international conference on document analysis and recognition*. IEEE, 2013, pp. 560–564.

[22] U.-V. Marti and H. Bunke, "The iam-database: an english sentence database for offline handwriting recognition," *International Journal on Document Analysis and Recognition*, vol. 5, no. 1, pp. 39–46, 2002.

[23] Y. Hannad, I. Siddiqi, and M. E. Y. El Kettani, "Arabic writer identification using local binary patterns (lbp) of handwritten fragments," in *Iberian conference on pattern recognition and image analysis*. Springer, 2015, pp. 237–244.

[24] M. Gargouri, S. Kanoun, and J.-M. Ogier, "Text-independent writer identification on online arabic handwriting," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, 2013, pp. 428–432.

[25] H. El Abed, V. Märgner, M. Kherallah, and A. M. Alimi, "Icdar 2009 online arabic handwriting recognition competition," in *2009 10th International Conference on Document Analysis and Recognition*. IEEE, 2009, pp. 1388–1392.

[26] T. Dhieb, W. Ouarda, H. Boubaker, and A. M. Alimi, "Beta-elliptic model for writer identification from online arabic handwriting." *Journal of Information Assurance & Security*, vol. 11, no. 5, 2016.

[27] H. El Abed, M. Kherallah, V. Märgner, and A. M. Alimi, "On-line arabic handwriting recognition competition," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 14, no. 1, pp. 15–23, 2011.

[28] R. Mitchell, J. Michalski, and T. Carbonell, *An artificial intelligence approach*. Springer, 2013.

[29] A. B. Hassanat, M. Alkasassbeh, M. Al-Awadi, and A. Esra'a, "Color-based object segmentation method using artificial neural network," *Simulation Modelling Practice and Theory*, vol. 64, pp. 3–17, 2016.

[30] T. Mitsa, *Temporal data mining*. Chapman and Hall/CRC, 2010.

[31] A. B. Hassanat, "Dimensionality invariant similarity measure," *arXiv preprint arXiv:1409.0923*, 2014.

[32] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[33] A. B. Hassanat, V. S. Prasath, M. Al-kasassbeh, A. S. Tarawneh, and A. J. Al-shamailh, "Magnetic energy-based feature extraction for low-quality fingerprint images," *Signal, Image and Video Processing*, vol. 12, no. 8, pp. 1471–1478, 2018.

[34] A. S. Tarawneh, C. Celik, A. B. Hassanat, and D. Chetverikov, "Detailed investigation of deep features with sparse representation and dimensionality reduction in cbir: A comparative study," *arXiv preprint arXiv:1811.09681*, 2018.

[35] A. B. Hassanat, "Furthest-pair-based binary search tree for speeding big data classification using k-nearest neighbors," *Big Data*, vol. 6, no. 3, pp. 225–235, 2018.

[36] A. B. Hassanat, "Two-point-based binary search trees for accelerating big data classification using knn," *PLoS ONE*, vol. 13, no. 11, p. e0207772, 2018.

[37] A. B. Hassanat, "Norm-based binary search trees for speeding up knn big data classification," *Computers*, vol. 7, no. 4, p. 54, 2018.