International Journal of Artificial Intelligence and Machine Learning
Volume 11 • Issue 2

A Survey on Arabic Handwritten Script Recognition Systems

Soumia Djaghbellou, Department of Computer Science, University of Mohammed El Bachir El Ibrahimi, Bordj Bou Arreridj, Algeria

Abderraouf Bouziane, LMSE Laboratory, Department of Computer Science, University of Mohammed El Bachir El Ibrahimi, Bordj Bou Arreridj, Algeria

Abdelouahab Attia, Independent Researcher, Algeria

https://orcid.org/0000-0003-1558-7273

Zahid Akhtar, Department of Computer Science, University of Memphis, USA

ABSTRACT

The optical character recognition (OCR) system is still an active research field in pattern recognition. Such systems can identify, recognize, and distinguish electronically between characters and texts, printed or handwritten. They can also do a transformation of such data type into machine-processable form to facilitate the interaction between user and machine in various applications. In this paper, the authors present the global structure of an OCR system, with its types (on-line and off-line), categories (printed and handwritten), and its main steps. They also focused on off-line handwritten Arabic character recognition and provided a list of the main datasets publicly available. This paper also presents a survey of the works that have been carried out over recent years. Finally, some open issues and potential research directions have been highlighted.

KEYWORDS

Handwritten Arabic Characters, OCR System, Off-Line Recognition, On-Line Recognition, Pattern Recognition

1. INTRODUCTION

The automatic text recognition, also known as optical character recognition (OCR), is a process in which the contents are transformed into comprehensible and machine-process-able representation for the purpose of archiving, conducting research, editing, reusing and transmitting the information. The outcomes of OCR systems can be used in various applications such as automatic check processing in banks, automatic mail sorting, filled form processing, postal code recognition, and writer/gender/personality identification. The OCR systems can be either "off-line" processing a piece of paper by optical scanning or "on-line" processing the text entered with a stylus on a touch/sensitive screens/surfaces, thus providing temporal information as well.

A huge number of researches have been carried out on recognizing various types of alphabet and handwritten characters in different languages such as Latin, Chinese, Japanese, Kanji, Hangul, Urdu and Persian. Comparatively very limited works have been done on Arabic text recognition or Arabic OCR. Research on Arabic OCR started in the 1970s (Al-Badr, 1995). The first published work in the field was in 1975 (Märgner, 2006). While, the first Arabic OCR system was available

DOI: 10.4018/IJAIML.20210701.oa9

in the 1990s (Cheriet, 2006). But, recognition of Arabic handwriting still faces challenges because of unique style and cursive handwritten characters (Abandah, 2010) (Sahlol, 2014). In fact, Arabic OCR system is very active and hot research topic, which is widely being explored in multidisciplinary settings. This paper presents a literature survey on Arabic handwritten script recognition systems considering global OCR structure, its types, content types and available datasets. Also, some future research directions and open issues are mentioned.

The organization of the paper is as follows. Section 2 details the origin and the main characteristics of the Arabic language and writing. Section 3 first presents a general structure of an AOCR system and then discusses the principal phases of an offline handwritten character recognition system. Section 4 is focused on a general comparison of previous works, approaches and databases. The future research and open issues are discussed in Section 6. Finally, Section 7 gives the conclusion of the paper.

2. THE ARABIC LANGUAGE AND SCRIPT, ORIGIN AND CHARACTERISTICS

The Arabic language (also Quranic and liturgical language) is being used by around 1.8 billion people. The estimated number of people speaking all varieties of Arabic language is as many as 422 million throughout the Middle East and northern Africa (Alginahi, 2013). It is celebrated as a global (universal) language on Dec 18 of each year. The Arabic language is famous for its beauty and diversity of its styles. It is a semi-cursive language in both forms (i.e., printed and handwritten). It is written from right to left. The Arabic writing and alphabet have many different characteristics, which make it unique, as also shown in Table 1.

Table 1. Arabic script characteristics

Description
Arabic alphabet consists of 28 basic letters with only 3 vowels (ا.ي.و), as shown in Fig. 1. These letters change their shape according to their position in a word. The notion of an uppercase or lowercase letter does not exist, so the writing is unicameral (Lawgali, 2015), for more details please see Table 2.
The Arabic writing is rich by diacritics and particularly with points. There are 15 letters among the 28 of the alphabets that include points above or under the letter with three points maximally. These points allow distinguishing pronunciation of Arabic letters, as is shown in Fig. 2.
Diacritic marks: In the Arabic language, vowels are not letters, but signs associated to letters and written over or under it (i.e., fatha, damma, sukun, madda and kasra). In addition, shadda used to show that the character is doubled and that the syllable is stressed (Alginahi, 2013), see Fig. 3. The Tanween (an n sound at the end of a word) represented with double fatha, double damma or double kasra (Alginahi, 2013), see Fig. 4. Hamza is another sign that can be considered as a special character or diacritic and can be written in different position (Alginahi, 2013), as shown in Fig. 5.
The Arabic characters can be joined either from right side, left side or from both sides. Only six of these characters (ا.د.ذ.ر.ز.و) are not connectable with their successors in a word, which causes a separation of the word into parts or sub-words (Jumari, 2002), as show Fig. 6.
In Arabic writing, the same character or the same word can be written with different styles/ways and sizes by different writers or even the same writer (Lawgali, 2015), as is shown in Fig. 7 and Fig. 8.

Figure 1. Arabic basic characters

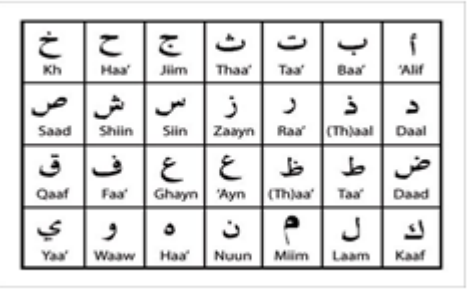


Figure 2. Positions of the points

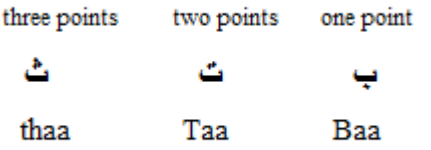


Figure 3. Diacritic marks

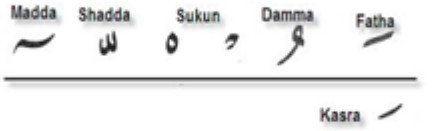


Figure 4. Tanween forms

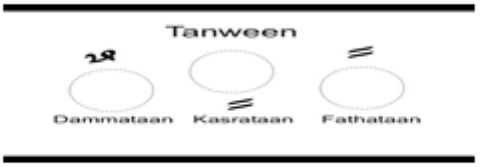


Figure 5. Positions of the Hamza

Figure 6. Sub words

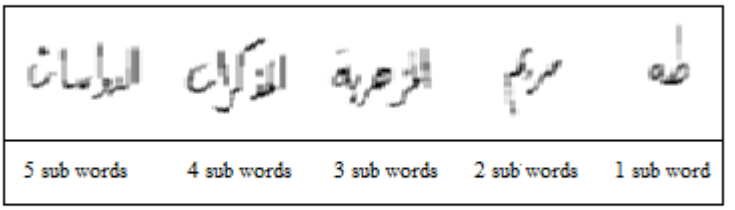


Figure 7. Characters written In different ways

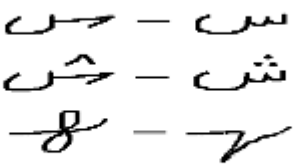
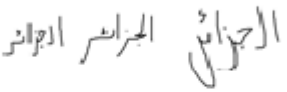


Figure 8. various writing styles of the same word



3. ARABIC OCR SYSTEMS

The Arabic optical character recognition systems aim at converting/translating automatically the images of Arabic handwritten or printed text into machine understandable words. There are two types of Arabic OCR system based on data acquisition method, i.e., on-line and off-line systems.

3.1 On-line Systems

On-line systems have ability to recognize a text or a character in real time employing special equipment such as a pen and a tablet (Jumari, 2002). In this type of systems, for further processing, handwritten characters in a particular script is classified and stored as Unicode or ASCII format. Online handwritten character recognition systems further can be divided into (1) writer dependent and (2) writer independent character recognition system (Kasturi, 2002).

3.2 Off-line Systems

This type of Arabic OCR systems is used in recognizing manuscripts from written/printed documents, the image of the written/printed text is scanned using a scanner. Please refer (Plamondon, 2000) for more details about the difference between these main two types systems. Further off-line systems can be of two modes as detailed below:

Table 2. Arabic alphabet shapes.

No	*Name*	*Isolated*	*beginning*	*middle*	*End*
1	Alif	ا	ا	ـا	ـا
2	Baa	ب	بـ	ـبـ	ـب
3	Taa	ت	تـ	ـتـ	ـت
4	Thaa	ث	ثـ	ـثـ	ـث
5	Jeem	ج	جـ	ـجـ	ـج
6	Haa	ح	حـ	ـحـ	ـح
7	Khaa	خ	خـ	ـخـ	ـخ
8	Daal	د	د	ـد	ـد
9	Thal	ذ	ذ	ـذ	ـذ
10	Raa	ر	ر	ـر	ـر
11	Zaa	ز	ز	ـز	ـز
12	Seen	س	سـ	ـسـ	ـس
13	Sheen	ش	شـ	ـشـ	ـش
14	Saad	ص	صـ	ـصـ	ـص
15	Dhad	ض	ضـ	ـضـ	ـض
16	Tta	ط	طـ	ـطـ	ـط
17	Dha	ظ	ظـ	ـظـ	ـظ
18	Ain	ع	عـ	ـعـ	ـع
19	Ghain	غ	غـ	ـغـ	ـغ
20	Faa	ف	فـ	ـفـ	ـف
21	Qaf	ق	قـ	ـقـ	ـق
22	Kaaf	ك	كـ	ـكـ	ـك
23	Lam	ل	لـ	ـلـ	ـل
24	Meem	م	مـ	ـمـ	ـم
25	Noon	ن	نـ	ـنـ	ـن
26	Haa	ه	هـ	ـهـ	ـه
27	Waaw	و	و	ـو	ـو
28	Yaa	ي	يـ	ـيـ	ـي

3.2.1 Printed Characters

They have one style and size for any given font. This mode of systems can be classified into two categories: single Font and multi font (Kasturi, 2002):

Single Font: In this category, the document is printed using a single font and hence this event makes it easy to recognize, as characteristics of the font remain same.

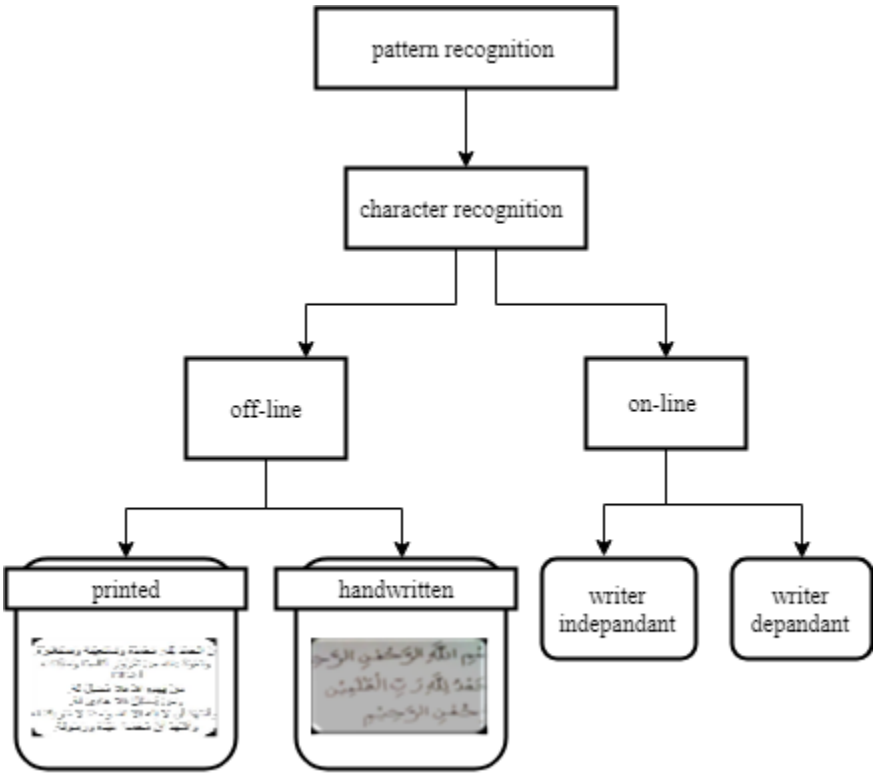
Multi Font: In this category, several fonts are used within the document.

3.2.2 Handwritten Characters

The system has to deal with different styles and sizes by the same writer as well as different writers (Lawgali, 2015). This category can be further divided into two groups based on how much content is stored in it, i.e., isolated character/symbol containing single character in an image, document containing group of lines, words and more characters.

In this survey article, the main interest is off-line handwritten recognition system, its types and modes, as shown in Fig. 9.

Figure 9. AOCR system, types and modes.

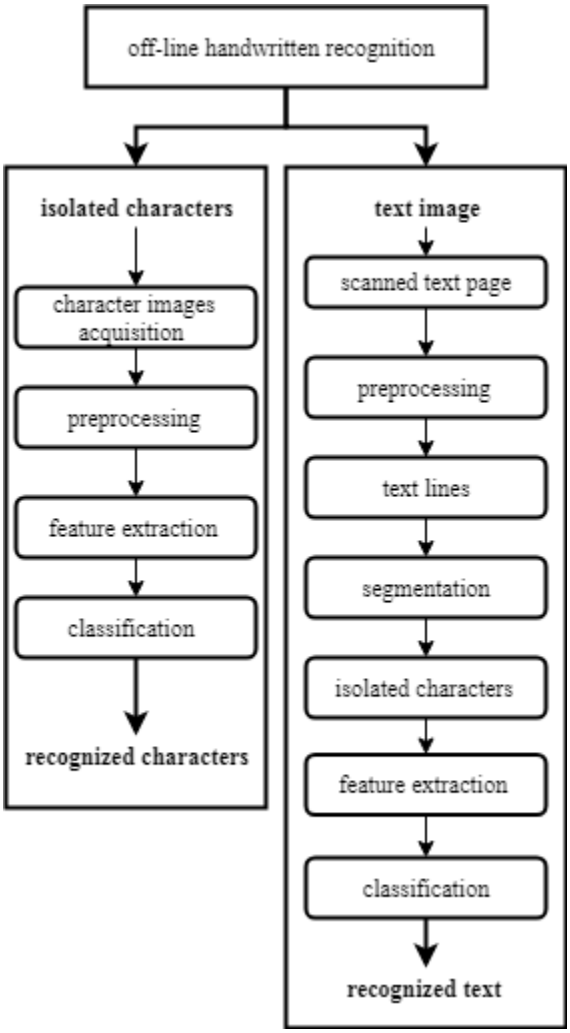


4. OFF-LINE HANDWRITTEN RECOGNITION SYSTEMS

An off-line handwritten recognition system needs several steps or phases to achieve a recognized text or character, starting by data acquisition, preprocessing operations, segmentation, extraction of the primitives and finishing by the classification stage, as is shown in Fig. 10.

Most of the Arabic off-line handwritten recognition research studies have been conducted using different standard databases that consist of alphabets, numbers, texts, images, special character. Each dataset is divided into two groups: training data and testing data. Table 3 shows the famous Arabic databases on Arabic handwritten characters, words, text and digits.

Figure 10. Off-line handwritten recognition system steps



4.1 Preprocessing Phase

After the image acquisition, the next important step in any recognition system is the preprocessing that includes different operations on the digitized image of a raw image to minimize noise and increase the capability of extracting features by cleaning and thinning the image. The operations are binarization (converting a grayscale image into bi-level image), smoothing (removing unwanted variations in the input image), thinning (minimizing the width of a line), normalization (scaling character to a fixed size and to center the character before recognition) and baseline detection (baseline means the line which contains information about the text orientation and the connection points between characters) (Jumari, 2002).

4.2 Segmentation Phase

After the preprocessing, the segmentation becomes an important stage, because it has an effect on the recognition rate (Lawgali, 2015). For segmentation, several techniques have been adopted to solve the problem of page decomposition, especially with the Arabic writing, its cursives and variation

Table 3. Summary of the famous Arabic datasets used in recognition systems

Database	*Data type*	***Authors***
AHCD	16800 isolated Arabic characters	(El-sawy el al. 2017)
AIA9K	Isolated Arabic handwritten alphabet (8737 letters with 28 classes)	(Torki el al. 2014)
HACDB	Arabic character images	(Lawgali et al. 2013)
SUST-ALT	Numerals datasets Letters datasets Arabic names datasets	(Musa et al. 2011)
APTI (Arabic Printed Text Images)	Words, word images and characters	(Slimane et al. 2009)
AD Base MAD Base	Digits	(El Sherif et al. 2007)
IFHCDB	Isolated characters and numerals	(Mozaffari el al. 2006)
IFN/ENIT	115585 pieces of Arabic words 26459 handwritten Tunisian town 212211 characters	(Pechwiz et al. 2002)
AHDB	10,000 words for check processing, written by 100 writers	(Al-Ma'adeed et al. 2002)
AI-ISRA	500 Arabic sentences 37000 Arabic words 10000 digits 2500 signatures written by 500 writers	(Kharma et al. 1999)
AHD/AMSH	12300 Arabic handwritten words written by 82 different writers.	(AL-NASSIRI et al. 2007)
KHATT	1000 forms, 2000 (random and fixed paragraphs) & free paragraphs written by 1000writers.	(Mahmoud et al. 2012)
CENPARMI	3,000 checks (Legal and courtesy amounts and digits).	(Al-Ohali et al. 2003)
Alamri Database	46,800 digits, 13,439 numerical strings, 21,426 letters, 11,375 words, 1,640 special symbols, written by 328 writers.	(Alamri et al. 2008)

in sizes and forms (Jumari, 2002). It is an operation of segmenting a page of text into three levels (Lawgali, 2015):

Decomposing a page into lines: Generally, a page or a text consists of several lines.

Segmenting a line into words: Depending on the longer spaces between/separating the words, the line is segmented into words, because most researchers assume in their technique that the space between words is bigger than the short space between sub-words (Kim, 1999).

Segmenting a word into individual characters: To get a set of individual characters, the word is segmented by detecting segmentation points, which identified the end of character and the beginning of the next one (Kim, 1999).

4.3 Feature Extraction Stage

Feature extraction is the central part of any recognition system, where the character produced in segmentation stage is used to extract some essential traits for the classification stage (Jumari, 2002). There are numerous types of features, but generally, they are classified into two types (Saeed, 2014):

4.3.1 Structural Features

To describe and extract local and global properties of the geometrical and topological characteristics suitable for the classification purposes, for an Arabic script, these features include loops, dots and their position, endpoints, branch points, strokes (width and height) (Lawgali,2015), as is shown in Fig. 1.

4.3.2 Statistical Features

They are numerical measures computed over images or regions of images, including pixel densities, histograms of chain code directions, moments and Fourier descriptors.

Fig. 12 depicts a simple example of this type of features that is text image dividing into zones and using the density of pixels in those zones as a feature (El Moubtahij, 2014).

Figure 11. Example of some structural features (loops, branch points, ends points).

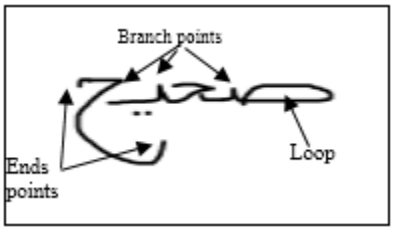
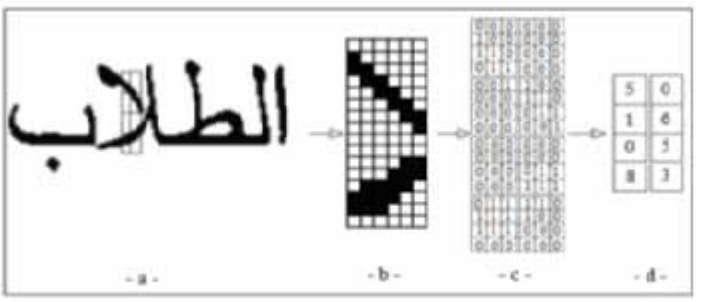


Figure 12. Intensity method for statistical features extraction.



4.4 Classification Process

The classification is the elaboration of a decision rule that transforms the attributes, extracted from the previous step, characterizing the forms (character) into a class membership (passage from the coding space to the decision space) (Safa, 2016).

Classification contains two main phases (safa, 2016):

Learning phase: This step builds a prototype dictionary. It is about grouping into class's several prototypes whose characteristics are getting closer. There are two types of learning: supervised and unsupervised, **where learning is said supervised, when the input data, entering the**

process, is already categorized and the algorithms must use it to predict a result in order to be able to do this operation later with another data not categorized. Unlike unsupervised learning, the data is communicated to the machine without providing it with the examples of results expected at the output.

Recognition and decision: The decision is the ultimate stage of recognition. From the parameter description of the processed character, the recognition module looks among the reference models present, which are closest to it, where it can lead to success if the answer is unique (only one model meets the description of the form of the character). It can lead to confusion if the answer is multiple (several models correspond to the description) (safa, 2016).

There are various classification methods **(classification algorithms or classifiers which use a set of features or parameters to characterize and categorize each object)** adopted by the researchers in many modern learning applications with many variations and combinations. The most adopted techniques are KNN (K nearest neighbors) (AlKhateeb, 2009), (Yousif, 2014), HMM (Hidden Markov Model) (AlKhateeb, 2011),(Lawgali, 2014), ANN (Artificial Neural Network) (Graves, 2009),(Elleuch, 2016) and SVM (Support Vector Machine) (Gazzah, 2008),(Khalifa, 2011).

4.5 Classification Performance Measures

Finally, to evaluate the performance of recognition or classification system, there are various measures being used:

Classification accuracy is the ratio of correct predictions to total predictions, which is multiplying by 100 to give a percentage result.

Error rate or misclassification rate equals to (1-(correct predictions /total predictions)) * 100.

Confusion matrix is a method for summarizing the performance of a classification algorithm. That can give a better idea of what the classification model is getting right and what types of errors it is making. Its content is organized into a table, or a matrix. Each row of the matrix corresponds to a predicted class. Each column of the matrix corresponds to an actual class.

5. LITERATURE ON ARABIC HANDWRITTEN RECOGNITION SYSTEMS

In this section, we present some notable existing works in the domain of Arabic handwritten recognition considering the method experimental analysis and dataset.

(Hani, 2016) developed a system using Histogram of Oriented Gradient (HOG) to extract feature vectors and the multi-class SVM with an RBF kernel for the classification. The used IFN/ENIT database, and the performance of their method was compared with Gabor filter. The presented HOG + SVM attained 1.51% (ECR) = 15 misclassified characters from 56 classes, and Gabor filter + SVM give 7.16% (ECR) = 68 misclassified character.

(Al-Jubouri, 2017) proposed a system for isolated Arabic handwritten characters based on two classifiers, i.e., support vector machine (SVM) and neural network (NN). They tested their system on IFN/ENIT database and achieved a total accuracy of 99.2%.

(El-Sawy, 2017) used the convolutional Neural network (CNN), which was trained and tested on the database collected by the authors. The dataset contains 16800 handwritten Arabic characters divided on training set with 13440 character images and testing set with 3360 character images. An average of 5.1% miss-classification error on testing data was achieved.

Similarly, (Younis, 2017) used, for the handwritten Arabic character recognition problem, a convolutional neural network (CNN) models with regularization parameters such as batch normalization to prevent over fitting. The AIA9K and AHCD databases were exploited to get, respectively, accuracies of 94.8% and 97.6%.

(Akram, 2017) designed a recognition system that is based on the hidden Markov Models (HMM) Toolkit (HTK) without explicit segmentation with the technique of sliding windows to extract a set

of features (features of local densities and features statistics). The simple database "Arabic-Numbers" and IFN/ENIT were employed to evaluate the performance of the proposed system, which achieved accuracy rates of 80.26% and 78.95%, respectively.

(Rabi, 2018) utilized Hidden Markov Models in the recognition system for cursive Arabic handwriting. The framework depends on the densities of the foreground pixels of the samples as features. For experiments, the benchmarking database IFN/ENIT was used to get a recognition rate of 87.93%.

(Abdulllah, 2018) modeled an OCR system for Arabic words from IFN/ENIT database using neural network classifier (NNC). Before feeding the input sample to NNC, a set of preprocessing steps were performed, i.e., removing spaces between words, baseline estimation, correction and resizing the words. The designed system gave a recognition rate of 70%.

(Jebril, 2018) proposed an Arabic character recognition system that is composed of three phases: i) the segmentation stage to extract characters, ii) use of Histograms of Oriented Gradient (HOG) for feature extraction, and iii) employment of SVM to classify characters. Experiments were performed on a dataset consisting of more than 43000 handwritten words (30000 for training and 13000 for testing). They reported a recognition rate of 99%.

(Assayony, 2018) presented a bag-of-features (BoF) based framework using Gabor filters (features and descriptors) to produce robust statistical features that can be utilized as a holistic handwritten word recognition system. For empirical analysis, a handwritten Arabic check's legal amount's public dataset was used. The best average recognition accuracy achieved by the proposed features was 86.44%.

(Rabi, 2018) developed a scheme for offline cursive Arabic handwritten texts recognition. The Hidden Markov Models (HMMs) after feature extraction stage without explicit segmentation was utilized. The experiments were done on benchmark IFN/ENIT database that procured a recognition rate of 87.93%.

(Hassan, 2019) presented an Arabic handwriting word recognition method, which uses scale invariant feature transform (SIFT) without and support vector machines (SVMs) in the classification step. A high recognition rate of 99.08% was reported on (AHDB) database.

(Ali, 2019) introduced a model that depends on fusion strategy for handwritten Arabic script recognition and identification of multi-font with SH Roqa, Naskh, Farsi, and Igaza. They tested their method on AHDB and AHCD datasets and achieved an excellent performance with higher accuracy.

(Salam, 2019) proposed an offline isolated Arabic handwriting character recognition system. It is a new architecture based on SVM with an authors' collected dataset for training and testing. The system achieved high recognition accuracy of 99.64%.

(Zanona, 2019) presented a handwritten Arabic characters recognition model based on a set of preprocessing functions with contour analysis to generate a vector that will be recognized using neural network architecture. They tested their system on a private data which includes all the 28 Arabic letters presented separately, to achieve a complete dataset accuracy of 98.8% and a complete dataset precision of 99.4%.

(Mohsin, 2020) designed a handwritten Arabic text recognition system which gets single lines from the text and convert them into words then into isolated characters, basing on a segmentation operation which constructs a histogram of comparison between pixels values and a threshold value, using a multilayer feed forward neural network for the classification process, to achieve a recognition rate of 83%.

(Sahlol, 2020) based on A hybrid machine learning approach with a binary whale optimization algorithm for feature extraction , trained and tested on benchmarking database CENPA-RMI to give results that show clear advantages of this proposed approach in terms of recognition accuracy= 96% and processor time(sec)=1.91.

Table 4. Summary of the previous presented works

Authors	*Recognition Technique*	*Database(s)*	*Performance*
(Hani 2016)	Arabic Handwritten Script Recognition System Based on HOG and Gabor Features	IFN/ENIT (words images)	Error classification rate (ECR)=1.51%
(Al-Jubouri, 2017)	Combining support vector machine (SVM) and neural network (NN) in the recognition system for isolated Arabic characters	IFN/ENIT	Recognition rate =99.2%
(El-Sawy, 2017)	Using convolutional neural network (CNN) on the database proposed for their recognition system.	New database proposed (16800 character images)	Misclassification error=5.1%
(Younis, 2017)	Deep neural network for the handwritten Arabic character recognition.	-AIA9K -AHCD	Recognition rate=94.8% Recognition rate=96.7%
(Akram, 2017)	Using hidden Markov Models (HMM), Toolkit (HTK) with sliding windows to extract features	IFN/ENIT	recognition rate =80.26% and 78.95%
(Rabi, 2018)	Hidden markov models for cursive Arabic handwriting recognition.	IFN/ENIT	Recognition rate= 87.93%
(Abdulllah, 2018)	NN classifier preceded by preprocessing techniques (removing spaces between words, baseline estimation, correction and resizing the words)	IFN/ENIT	Recognition rate =70%
(Jebril, 2018)	Using Histogram of oriented gradient (HOG) for feature extraction and SVM for classification.	Dataset of 43000 handwritten words	Recognition rate=99%
(Assayony, 2018)	Bag-of-features (BoF) framework based on Gabor filters (features and descriptors) for producing robust statistical features for the recognition system	handwritten Arabic checks legal amounts public dataset	Recognition rate=86.44%
(Rabi, 2018)	Offline recognition of cursive Arabic handwritten text based on Hidden Markov Models (HMMs) without explicit segmentation.	IFN/ENIT	Recognition rate=87.93%
(Hassan, 2019)	Arabic Handwriting Word Recognition without segmentation and Based on Scale Invariant Feature Transform and Support Vector Machines (SVMs).	AHDB	Recognition rate = 99.08%
(Ali, 2019)	Arabic Handwriting Word Recognition without segmentation and Based on Scale Invariant Feature Transform and Support Vector Machines (SVMs).	AHDB	Recognition rate = 99.08%
(Salam, 2019)	New architecture based on SVM	Private (proposed) dataset	Recognition rate= 99.63%
(Zanona,2019)	Neural network architecture based on a set of preprocessing operation.	Private dataset(individual Arabic characters)	Average accuracy=98.8% Average precision=99.4%

continued on next page

Table 4. Continued

Authors	*Recognition Technique*	*Database(s)*	*Performance*
(Mohsin, 2020)	Handwritten Arabic text/ character recognition model based on a segmentation process and multilayer feed forward neural network as classifier.	Proposed dataset (images of handwritten Arabic text/lines)	Recognition rate= 83%
(Sahlol,2020)	A hybrid machine learning approach with a binary whale optimization algorithm for feature extraction.	CENPA-RMI	Accuracy= 96% Time (sec)= 1.91

6. OPEN ISSUES AND FUTURE RESEARCH DIRECTIONS

6.1. Generalization Ability:

Most existing systems perform well on given dataset for which they were developed, but their performance drops when tested on different dataset. **So for this reason it is important to experiment each developed system on more than dataset in order to improve its performance and to be able to trait other type of dataset content .**

6.2. Use of Deep Learning:

Despite recent remarkable accuracy, relatively very less frameworks have been developed using deep learning for Arabic handwritten character recognition, **this learning architecture has its own design and components that allow a deeper recognition process and best results in contrast with the complexity of Arabic writing.**

6.3. Taxonomy:

There is lack of standard taxonomy for Arabic OCR system. It is important to put the correct taxonomy that will help to obtain quality research works.

6.4. Large scale evaluation:

There is also lack of wide-reaching evaluation of prior works on Arabic OCR systems. Any such study will help to see the true progress in the field.

6.5. Reproducible Research:

Reproducible research should be encouraged in the field. It will hugely help to study scalability, **to show actual applications and to give to the researchers the ability of working and developing the existing studies done on recognition systems, to get more improved results.**

7. CONCLUSION

This paper presented the main characteristics of Arabic script OCR recognition systems as well as illustrating the global architecture of an Arabic character recognition system with their types and modes focusing on initial phases. Also, we presented the well-known available databases used for Arabic handwritten character recognition systems. We then reviewed some existing recent works in the field, a comparative summary is provided with their various techniques, datasets, complexity and nature. Some open issues and future research directions show that it is a challenging task that needs more experiments and studies.

REFERENCES

Abandah, G. A., & Malas, T. M. (2010). Feature selection for recognizing handwritten Arabic letters. *Dirasat Engineering Sciences Journal, 37*(2).

Akram, H., & Khalid, S. (2017). Using features of local densities, statistics and HMM toolkit (HTK) for offline Arabic handwriting text recognition. *Journal of Electrical Systems and Information Technology*, *4*(3), 387–396.

Al-Badr, B., & Mahmoud, S. A. (1995). Survey and bibliography of Arabic optical text recognition. *Signal Processing*, *41*(1), 49–77. doi:10.1016/0165-1684(94)00090-M

Al-Jubouri, M. A. H. (2017). Offline Arabic Handwritten Isolated Character Recognition System Using Support vector Machine and Neural Network. *Journal of Theoretical & Applied Information Technology*, *95*(10).

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.

Al-Nassiri, A. M. E. R., & Abdulla, S. A. (n.d.). *A new Arabic (AHD/AMSH) handwritten database*. Academic Press.

Al-Ohali, Y., Cheriet, M., & Suen, C. (2003). Databases for recognition of handwritten Arabic cheques. *Pattern Recognition*, *36*(1), 111–121.

Alamri, H., Sadri, J., Suen, C. Y., & Nobile, N. (2008). A novel comprehensive database for Arabic off-line handwriting recognition. In *Proceedings of 11th International Conference on Frontiers in Handwriting Recognition, ICFHR* (Vol. 8, pp. 664-669). Academic Press.

Alginahi, Y. M. (2013). A survey on Arabic character segmentation. *International Journal on Document Analysis and Recognition*, *16*(2), 105–126.

Ali, A. A. A., & Suresha, M. (2019). A new design based-fusion of features to recognize Arabic handwritten characters. *International Journal of Engineering and Advanced Technology*, *8*(5), 2570–2574.

AlKhateeb, J. H., Khelifi, F., Jiang, J., & Ipson, S. S. (2009, November). A new approach for off-line handwritten Arabic word recognition using KNN classifier. In *2009 IEEE International Conference on Signal and Image Processing Applications* (pp. 191-194). IEEE.

AlKhateeb, J. H., Ren, J., Jiang, J., & Al-Muhtaseb, H. (2011). Offline handwritten Arabic cursive text recognition using Hidden Markov Models and re-ranking. *Pattern Recognition Letters*, *32*(8), 1081–1088.

Assayony, M. O., & Mahmoud, S. A. (2018). Recognition of Arabic handwritten words using Gabor-based bag-of-features framework. *International Journal of Computing and Digital Systems*, *7*(01), 35–42.

Cheriet, M. (2006, September). Visual recognition of Arabic handwriting: challenges and new directions. In *Summit on Arabic and Chinese Handwriting Recognition* (pp. 1–21). Springer.

El Moubtahij, H., Halli, A., & Satori, K. (2014). Review of feature extraction techniques for offline handwriting arabic text recognition. *International Journal of Advances in Engineering and Technology*, *7*(1), 50.

El-Sawy, A., Loey, M., & El-Bakry, H. (2017). Arabic Handwritten Characters Recognition Using Convolutional Neural Network. *WSEAS Transactions on Computer Research*, *5*, 11–19.

El-Sawy, A., Loey, M., & Hazem, E. B. (2017). Arabic handwritten characters recognition using convolutional neural network. *WSEAS Transactions on Computer Research*, *5*, 11–19.

El-Sherif, E. A., & Abdelazeem, S. (2007, July). A Two-Stage System for Arabic Handwritten Digit Recognition Tested on a New Large Database. In Artificial intelligence and pattern recognition (pp. 237-242). Academic Press.

Elleuch, M., Maalej, R., & Kherallah, M. (2016). A new design based-SVM of the CNN classifier architecture with dropout for offline Arabic handwritten recognition. *Procedia Computer Science*, *80*, 1712–1723.

Gazzah, S., & Amara, N. B. (2008). Neural networks and support vector machines classifiers for writer identification using Arabic script. *The International Arab Journal of Information Technology*, *5*(1).

Graves, A., & Schmidhuber, J. (2009). Offline handwriting recognition with multidimensional recurrent neural networks. In Advances in neural information processing systems (pp. 545-552). Academic Press.

Hani, A., Elleuch, M., & Kherallah, M.(2016). *Arabic Handwritten Script Recognition System Based on HOG and Gabor Features*. Academic Press.

Hassan, A. K. A., Mahdi, B. S., & Mohammed, A. A. (2019). Arabic handwriting word recognition based on scale invariant feature transform and support vector machine. *Iraqi Journal of Science*, 381-387.

Jebril, N. A., Al-Zoubi, H. R., & Al-Haija, Q. A. (2018). Recognition of handwritten arabic characters using histograms of oriented gradient (HOG). *Pattern Recognition and Image Analysis*, *28*(2), 321–345.

Jumari, K., & Ali, M. A. (2002). A survey and comparative evaluation of selected off-line Arabic handwritten character recognition systems. *Jurnal Teknologi*, *36*(1), 1–18.

Kasturi, R., O'gorman, L., & Govindaraju, V. (2002). Document image analysis: A primer. *Sadhana*, *27*(1), 3–22.

Khalifa, M., & BingRu, Y. (2011, April). A novel word based arabic handwritten recognition system using svm classifier. In *International Conference on Electronic Commerce, Web Application, and Communication* (pp. 163-171). Springer.

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.

Kim, G., Govindaraju, V., & Srihari, S. N. (1999). An architecture for handwritten text recognition systems. *International Journal on Document Analysis and Recognition*, *2*(1), 37–44.

Lawgali, A. (2014). *An evaluation of methods for Arabic character recognition*. Academic Press.

Lawgali, A. (2015). *A survey on Arabic character recognition*. Academic Press.

Lawgali, A., Angelova, M., & Bouridane, A. (2013, June). HACDB: Handwritten Arabic characters database for automatic character recognition. In *European Workshop on Visual Information Processing (EUVIP)* (pp. 255-259). IEEE.

Mahmoud, S. A., Ahmad, I., Al-Khatib, W. G., Alshayeb, M., Parvez, M. T., Märgner, V., & Fink, G. A. (2012). KHATT: An open Arabic offline handwritten text database. *Pattern Recognition*, *47*(3), 1096–1112.

Manal, A., Afnan, A., & Mariam, A. (2018). Arabic handwriting recognition using neural network classifier. *Journal of Fundamental and Applied Sciences*, *10*(no 4S), 265–270.

Märgner, V., & El Abed, H. (2006, September). Databases and competitions: strategies to improve Arabic recognition systems. In *Summit on Arabic and Chinese Handwriting Recognition* (pp. 82–103). Springer.

Mohsin, A., & Sadoon, M. (2020). Developing an Arabic Handwritten Recognition System by Means of Artificial Neural Network. *Journal of Engineering and Applied Sciences, 15*(1), 1-3.

Mozaffari, S., Faez, K., Faradji, F., Ziaratban, M., & Golzan, S. M. (2006, October). *A comprehensive isolated Farsi/Arabic character database for handwritten OCR research*. Academic Press.

Musa, M. E. (2011, October). Arabic handwritten datasets for pattern recognition and machine learning. In *2011 5th International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1-3). IEEE.

Pechwitz, M., Maddouri, S. S., Märgner, V., Ellouze, N., & Amiri, H. (2002, October). IFN/ENIT-database of handwritten Arabic words. In *Proc. of CIFED* (*Vol. 2*, pp. 127-136). Citeseer.

Plamondon, R., & Srihari, S. N. (2000). Online and off-line handwriting recognition: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 63–84.

Rabi, M., Amrouch, M., & Mahani, Z. (2018). *Cursive Arabic Handwriting Recognition System Without Explicit Segmentation Based on Hidden Markov Models*. Academic Press.

Rabi, M., Amrouch, M., & Mahani, Z. (2018). Recognition of cursive Arabic handwritten text using embedded training based on hidden Markov models. *International Journal of Pattern Recognition and Artificial Intelligence*, *32*(01), 1860007.

Saeed, U. (2014). Automatic recognition of handwritten Arabic text: A survey. *Life Science Journal*, *11*(3s).

Safa, A. B. A., & Chikh, M. (2016). *Intitulé Reconnaissance Des Mots Arabes Manuscrites* (Doctoral dissertation).

Sahlol, A., & Suen, C. (2014). *A novel method for the recognition of isolated handwritten Arabic characters*. arXiv preprint arXiv:1402.6650.

Sahlol, A. T., Abd Elaziz, M., Al-Qaness, M. A., & Kim, S. (2020). Handwritten Arabic Optical Character Recognition Approach Based on Hybrid Whale Optimization Algorithm With Neighborhood Rough Set. *IEEE Access: Practical Innovations, Open Solutions*, *8*, 23011–23021.

Salam, M., & Hassan, A. A. (2019). Offline isolated arabic handwriting character recognition system based on SVM. *The International Arab Journal of Information Technology*, *16*(3), 467–472.

Slimane, F., Ingold, R., Kanoun, S., Alimi, A. M., & Hennebert, J. (2009, July). A new arabic printed text image database and evaluation protocols. In *2009 10th International Conference on Document Analysis and Recognition* (pp. 946-950). IEEE.

Torki, M., Hussein, M. E., Elsallamy, A., Fayyaz, M., & Yaser, S. (2014). *Window-based descriptors for arabic handwritten alphabet recognition: A comparative study on a novel dataset*. arXiv preprint arXiv:1411.3519.

Younis, K. S. (2017). Arabic handwritten character recognition based on deep convolutional neural networks. *Jordanian Journal of Computers and Information Technology*, *3*(3), 186–200.

Yousif, I., & Shaout, A. (2014). Off-Line handwriting Arabic text recognition: A survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, *4*(9).

Zanona, M. A., Abuhamdah, A., & El-Zaghmouri, B. M. (2019). Arabic Hand Written Character Recognition Based on Contour Matching and Neural Network. *Computer and Information Science, 12*(2).