Handwritten Arabic Text Recognition using Principal Component Analysis and Support Vector Machines

Faisal Al-Saqqar[1], Mofleh Al-Diabat[3], Mesbah Aloun[4]
Department of Computer Science
Al al-Bayt University, Mafraq, Jordan

Atallah. M AL-Shatnawi[2]
Department of Information Systems
Al al-Bayt University, Mafraq, Jordan

***Abstract*—In this paper, an offline holistic handwritten Arabic text recognition system based on Principal Component Analysis (PCA) and Support Vector Machine (SVM) classifiers is proposed. The proposed system consists of three primary stages: preliminary processing, feature extraction using PCA, and classification using the polynomial, linear, and Gaussian SVM classifiers. In this proposed system, text skeleton is first extracted and the images of the text are normalized into uniform size for extraction of the global features of the Arabic words using PCA. Recognition performance of this proposed system was evaluated on version 2 of the IFN/ENIT database of handwritten Arabic text using the polynomial, linear, and Gaussian SVM classifiers. The classification results of the proposed system were compared with the results produced by a benchmark. TRS that is depending on the Discrete Cosine Transform (DCT) method using numerous normalization sizes of Arabic text images. The experimental testing results support the effectiveness of the proposed system in holistic recognition of the handwritten Arabic text.**

***Keywords*—Handwritten Arabic text; holistic recognition; principal component analysis; support vector machines**

I. Introduction

The ultimate objective of any Arabic Text Recognition System (ATRS) is to imitate the human understanding abilities so that the computer can read, understand, and accomplish activities on texts that are similar to the ones which the human mind executes [2, 3, 4, 7, 26]. The Arabic language is universal language and the official language of 25 countries and greater than 300 million individuals in the world. Additionally, many Arabic characters are utilized in numerous languages such the Iranian, Jawi, and Urdu languages [1, 7, 3]. Review of the literature uncovers that, so far, there are two major systems for offline Arabic text recognition; segmentation-free systems (holistic recognition approaches) and segmentation-based systems [25, 12]. In the former systems, recognition is applied on the entire representation of the text or word, which is treated as one unit with no segmentation. In the segmentation-based systems, however, cursive text is often segmented into characters or small segments called primitives. This approach, thus, suffers from varying problems, including overlapping and ligatures, short distances between connected characters, and the Arabic writing properties [6]. Details on characteristics of the handwritten Arabic text can be found in Al-Shatnawi et al. [6] and Al-Shatnawi [4]. For the holistic approaches in text recognition, a universal feature vector is computed for the indivisible input texts or words for them to be classified by using any of a number of machine learning approaches [13].

The ultimate objective of feature extraction is to produce efficient representation of the image of the text using a group of distinctive characteristics. These characteristics may be categorized into three classes: (i) high-level characteristics, which are drawn from the entire image of the text or word, (ii) medium-level characteristics that are derived from the characters, and (iii) low-level characteristics, which are usually extracted from the related sub-characters [18]. In other respects, the handwritten Arabic text may be recognized using various classifiers like the Support Vector Machines (SVM), Hidden Markov Model (HMM), the k-nearest neighbors (kNN), and the Artificial Neural Network (ANN) classifiers [11, 12, 26].

This study proposes a multi-stage Offline Holistic Handwritten Arabic Text Recognition System (OHATRS) based on Principal Component Analysis (PCA) and SVM classifiers. This suggested system progresses in three steps: preliminary preprocessing, feature extraction using PCA, and classification using the polynomial, linear, and Gaussian SVM classifiers. The primary contributions of this paper can be abbreviated as follows: (i) extracting the statistical handwritten Arabic text features using the PCA technique, (ii) testing and evaluating the extracted features on version 2 of the IFN/ENIT database of handwritten Arabic text using the polynomial, linear, and Gaussian SVM classifiers, and (iii) comparing the recognition results of the proposed OHATRS with benchmark ATRS that is depending on the Discrete Cosine Transform (DCT) method.

The remainder of this paper is organized as follows: Section 2 overviews previous holistic handwritten ATRSs while Section 3 presents the proposed OHATRS. Thereafter, Section 4 presents the experimental recognition results of the herein proposed system and discusses them. Then, Section 5 outlines the conclusions of this study and highlights directions for future research.

II. Related Work

Several previous research efforts have examined the offline holistic approaches to recognizing the handwritten Arabic cursive scripts. For instance, El-Bashir [21] suggested recognition of Arabic sub-words using PCA as the means of feature extraction. Recognition was carried in his study by using different norms that are, ENorm and EEuclidean norm. The suggested system was verified on dataset of two groups, one comprising two-character sub-words and the other including three-character sub-words. The evaluation results exposed that the classification accuracies related with the first norm, ENorm, second norm, and EEuclidean were 74.3%, 76.8%, 76.8%, and 77.17%, respectively, in the instance of the two-character sub-words and 75.85%, 77.45%, 78.2%, and 78.49%, respectively, in the instance of the three-character sub-words.

Sagheer et al. [27] suggested holistic recognition model for the handwritten Urdu words depending on sets of integrated features and the SVM classifier. Their suggested system incorporated the gradient, or directional, features and the structural features that were extracted from the handwritten Urdu words. When tested on the CENPARMI Urdu Words Database, this proposed system achieved a recognition accuracy of 97.00%.

Nemmour and Chibani [15] presented an offline holistic model for recognition of the handwritten Arabic text based on combination of the SVMs and Ridgelet transform. The Ridgelets were employed to generate relevant features of handwritten words whereas classification was based on the 'One-Against-All' multi-class operation of the SVMs. This system was then tested on vocabulary of 24 words taken from the IFN/ENIT database. Ridgelet performance was evaluated in this study relative to the results produced by the Radon and uniform grid (zoning) feature method. The performance evaluation outcomes spotlight reliability of the combination of the SVM and Ridgelet tools for recognition of the handwritten Arabic words.

EI Qacimy et al. [11] suggested offline, word-based system for recognition of the handwritten Arabic text depending on the DCT features and a SVM classifier that is improved by reject option. This system comprised four key processes, namely, preprocessing, segmentation into sub-words, extraction based on DCT features, and classification by the SVM RBF classifier. The system was then evaluated on 2,000 word images that were chosen randomly from the IFN/ENIT database of handwritten Arabic words that were separated into a training sub-set of 1,500 images and a testing sub-set of 500 images. Afterwards, performance of this suggested system was verified with the ranks of performance of state-of-the-art schemes that used DCT features for classification of the Arabic handwritten text. The results disclosed effectiveness of this proposed system in holistic classification of the Arabic words.

Kadhm and Abdul [24] developed an offline holistic system for recognition of the handwritten Arabic words that is based on SVM classifiers. The HOG and DCT were both employed for feature extraction. Then, this system was verified on the AHDB Database, which contains 2,913 images of handwritten Arabic words by the SVM linear, polynomial, and RBF kernel classifiers. These three classifiers produced recognition accuracies of 96.32%, 92.63%, and 91.50%, respectively.

Hassan and Alawi [9] developed a holistic, offline, system for recognition of the handwritten Arabic words depending on the SVM with the Gaussian kernel and the Discrete Wavelet Transform (DWT) transforms. This system was established based on four levels of the DWT by segmenting the wavelet space into 16x16 blocks. Then, the standard deviation was calculated for every block. Performance of this system was then evaluated on database of 1,160 word images of names of Iraqi cities that had been handwritten by 30 writers of differing educational backgrounds and ages using the SVM Gaussian, polynomial, and linear kernel classifiers, which produced recognition accuracies of 89.17%, 90.00%, and 90.65%, respectively.

III. The Proposed Text Recognition System

This paper presents a multi-stage system for holistic recognition of the handwritten Arabic text that is based on PCA and SVM classifiers. The proposed (OHATRS) has three basic processes: preliminary preprocessing, feature extraction using PCA, and classification using the Gaussian, linear, and polynomial SVM classifiers. Architecture of this proposed system is presented in Fig. 1. In the preliminary stage, text skeleton is first extracted and images of the text are then normalized into uniform size for the purpose of extraction of the universal features of the Arabic text using PCA. The extracted features are then used to classify the handwritten Arabic text using the aforementioned SVM classifiers. Stages of the OHATRS are illustrated in the following sub-sections.

A. *Preliminary Stage*

The preliminary processing stage prepares the text data under consideration for the successive stages. It enhances uniformity of the texts, which is an essential requirement of the recognition system. Preliminary processing is concerned with representation of the Arabic text images and normalization processes. At this stage, the skeleton of the word is first extracted by means of the skeletonization-based morphological process so as to eliminate the unnecessary pixels via extraction of the text skeleton at the single-pixel width level. Afterwards, the image of the text of concern is normalized into suitable size for extraction of the global features of the Arabic text using the PCA technique. A briefing of the two operations making up this stage follows.

B. *The Skeleton Extraction Process*

In the skeleton extraction process, skeleton of the input text image is extracted via the skeletonization-based morphological operation method. As such, this process refines text shape and minimizes the size of the data that needs handling for the purpose of feature extraction and recognition [5]. This particular approach was selected to thin the handwritten Arabic text because it proved (e.g., [2]) to be having high performance in thinning the handwritten Arabic text. An example on skeletons of handwritten Arabic texts that have been extracted using the skeletonization-based morphological method is presented in Fig. 2.

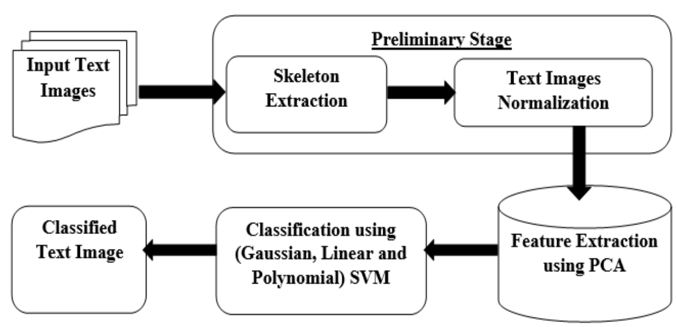


Fig. 1. Architecture of the Proposed OHATRS.

(a) Original Image. (b) Produced Skeleton.

Fig. 2. Example of (a) Handwritten Arabic Text before thinning and (b) Extracted Skeleton of this Text.

C. Normalization

Normalization of the text images is a very important step in the text recognition process. Because the styles of writing differ from one person to another, the size normalization process is often employed to convert the sizes of the characters or words into a uniform standard size [16]. In view of importance of this process and its effect on the recognition results of the system proposed in this paper (OHATRS), recognition performance of this system was tested on text images of varying sizes, taking into consideration the smallest and largest image sizes in the relating databases so as to select the best results and compare performance of this proposed system with levels of performance of another recognition system. Those sizes and their effects on performance of the proposed system are discussed in the experimental results and discussion section.

D. Feature Extraction using Principal Component Analysis

Feature extraction is the most important process in the recognition systems of the handwritten texts. Best recognition usually depends on successful use of efficient feature extraction methods [9]. The eventual goal of feature extraction is to produce efficient representation of the entire text image through set of features [18].

Principal Component Analysis (PCA) is a statistical linear transform technique. It was developed originally by Pearson [20]. It is broadly employed for differing pattern recognition applications like character recognition (e.g., Abandah et al. [14]), data compression (e.g., da Rocha Gesualdi and Seixas [8]), and face recognition (e.g., Bansal et al. [17]).

This study applied PCA to extract and select the relevant features of handwritten Arabic text as a global statistical text feature extraction technique for the extracted features to be classified by the SVM classifiers. The PCA is commonly employed as feature extraction method so as to reduce dimensions of images to manageable sizes. PCA begins by calculating the mean of the data matrix. Then, it computes the covariance of the data. Thereafter, the Eigenvalues and Eigenvectors are estimated [28]. The PCA aims at finding the space that represents direction of the maximal variance of the data under consideration. It defines low-dimensional space, or a PCA space (W), that can be employed to transform the data (X = {x1, x2, . . . , xn}, where n is number of observations or samples and xi is the ith observation, sample, or pattern) from high-dimensional space into low-dimensional space [10].

Principal Component Analysis has been already applied successfully in feature extraction and in dimension reduction in numerous recognition systems of isolated Arabic characters. For example, Ali and Shaout [28] employed PCA for feature extraction in recognition of isolated, handwritten, Arabic characters using the Adaptive Neural Network Fuzzy Inference System (ANFIS). As well, Khan et al. [19] employed PCA in recognition of isolated Urdu characters. In addition, Abandah et al. [14] employed PCA for reduction of the dimensionality of the features extracted from isolated Arabic letters (characters) for text recognition purposes by using five classifiers: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Diagonal Quadratic Discriminant Analysis (DQDA), k-NN, and Diagonal Linear Discriminant Analysis (DLDA).

In this process, the PCA is applied to extract and select global Arabic text features according to the following six steps [10]:

Step 1: The two-dimensional (2D) text image is transformed into mono-dimensional vector by concatenating every column (or row) in the 2D matrix in order to create long vector. If we have an M vector of the size N that represents sample image, then

$$X_i = [p_1 \ldots p_N]^T, i = 1, \ldots, N \qquad (1)$$

where X_i is a vector, px is pixel value Xi, and T is tanspose of the vector set.

Step 2: Find mean, μ, of the image, which can be calculated as follows:

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i \qquad (2)$$

Step 3: Find the mean center of the image, wi:

$$w_i = X_i - \mu \qquad (3)$$

Step 4: Calculate the covariance matrix, S. This matrix measures the relations between two dimensions or more. It can be computed from the equation:

$$S = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)(x_i - \mu)^{T_i} \qquad (4)$$

Step 5: Calculate the Eigenvalue, λ, and Eigenvector, V, of S according to the equation

$$SV_i = \lambda_i V_i, for\ i = 1,2 \ldots n. \qquad (5)$$

Afterwards, sort the eigenvectors based on their concomitant eigenvalues.

Step 6: Choose the eigenvectors which have the highest eigenvalues, W = {v1. . . vk}. The chosen W values correspond to the projection space of the PCA. Thereafter, project those W values on the low-dimensional space of the PCA.

E. Classification using SVM Classifiers

The SVM is a relatively modern classifier that employs kernels to find the optimum decision boundary and, then, separate between the potential classes in the high-dimensional feature spaces. Algorithm of the SVM was introduced originally by Vapnik [30]. It was proposed initially for the binary separation problems. However, it may be generalized easily to the multi-class classification problems. The fundamental form of the linear SVM classifier attempts to discovery the optimum hyperplane that separate the best set of samples that belong to differing classes [11].

In classification in the present study, the multi-class polynomial, Gaussian, and linear SVM classifiers were used for classification of images of handwritten Arabic text by using the sequent SVM kernel functions [22]:

The linear function: K(x, y) = (K(xi, xj) = (xixj))

The polynomial function: K(x, y) =(K(xi, xj)=(γxixj + coef)

The Gaussian function: K(x, y) = exp−γ||xi−xj ||2)

The proposed (OHATRS), which is based on the PCA and the Gaussian, linear, and polynomial SVM classifiers is presented in the following algorithm:

Algorithm (1): The proposed system for holistic recognition of the handwritten Arabic text based on PCA and the Gaussian, linear, and polynomial SVM classifiers.

Input: Images of Handwritten Arabic Text

Output: Classified Word

{Read the image of the handwritten Arabic text,
prepare the data of the text image using the following two preliminary steps:
- Extract the text skeleton using the skeletonization-based morphological operation method,
- Normalize the text image size to a suitable size,

Extract the global text features using PCA,
and then classify the text images using the Gaussian, linear, and polynomial SVM classifiers,
end}

IV. Experimental Results and Discussion

In this study, both the proposed and benchmark Arabic text recognition systems were implemented in the MATLAB 2017a environment in personal computer with an i3 processor, a speed of 1.90 GHz, and a memory of 6 GB. The various systems under study were tested on version 2.0 of the IFN/ENIT Arabic handwritten text database. This database is made up of 32,492 images of handwritten Arabic names of Tunisian villages and towns. The names are categorized into five sub-sets; a, b, c, d, and e sub-sets [23, 29]. The (a), (b), (c), and (d) sub-sets were employed for training purposes whereas the (d) and (e) sub-sets were utilized for testing purposes. All sub-sets are provided with ground truth information that has been employed in labeling the recognition observations.

For the purpose of verifying performance of the herein proposed OHATRS, the recognition results of the proposed system were compared with those of a benchmark recognition system that is based on DCT. The DCT method was employed by EI Qacimy et al. [11] for holistic classification of Arabic texts with the reject option on the basis of sub-word segmentation.

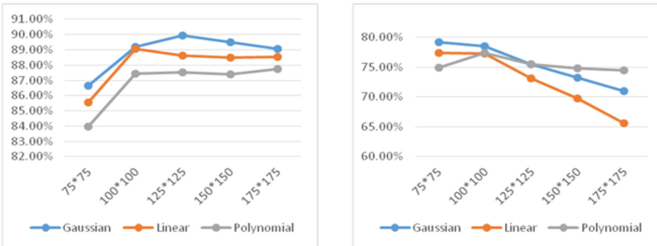
The OHATRS proposed here and the benchmark ATRS were tested on the (d) and (e) sub-sets of the IFN/ENIT database using (i) the polynomial, Gaussian, and linear SVM classifiers, and (ii) five normalized word image sizes: 75x75, 100x100, 125x125, 150x150, and 175x175. The cassification accuracies of both systems when tested on sub-set (d) of the IFN/ENIT database are summarized by Table I.

As Table I shows, the proposed OHATRS produced better classification accuracies than the benchmark ATRS system when applying the aforementioned five normalized image sizes on sub-set (d) of the IFN/ENIT database and using the Gaussian (RBF), linear, and polynomial SVM classifiers. The recognition accuracies of the proposed and the benchmark systems are shown in Fig. 3. The best classification accuracy (89.96%) achieved by the proposed OHATRS was concomitant with 125x125 image normalization size and the Gaussian SVM classifier. On the other hand, the best classification accuracy (79.14%) achieved by the benchmark ATRS was achieved with the 75x75 image normalization size and the Gaussian SVM classifier, too. This finding supports effectiveness of the proposed OHATRS in holistic recognition of the handwritten Arabic text. Table II shows the classification accuracies of both the proposed OHATRS and the benchmark ATRS when tested on sub-set (e) of the IFN/ENIT database.

Table II uncovers that better classification accuracies are associated with the proposed OHATRS than with the benchmark ATRS system when using (i) the same five normalized image sizes on sub-set (e) of the IFN/ENIT database and (ii) the Gaussian (RBF), linear, and polynomial SVM classifiers. The recognition accuracies of both systems are presented in Fig. 4. The best classification accuracy (77.80%) produced by the proposed OHATRS was associated with the 125x125 image normalization size and the Gaussian SVM classifier. Meanwhile, the best classification accuracy (68.46%) generated by the benchmark ATRS was concomitant to the 100x100 image normalization size and the polynomial SVM classifier. This result suggests effectiveness of the proposed OHATRS in holistic recognition of the handwritten Arabic text.

TABLE. I. The Classification Accuracies of the Proposed OHATRS and the Benchmark ATRS when Tested on Set (d) of the IFN/ENIT Database using the SVM Classifiers

	Normalization size	Classification accuracy		
		Gaussian	Linear	Polynomial
Proposed system based on PCA	75x75	86.64%	83.99%	85.54%
	100x100	89.18%	87.42%	89.07%
	125x125	89.96%	87.53%	88.62%
	150 x150	89.51%	87.41%	88.51%
	175x175	89.07%	87.75%	88.52%
Benchmark system using the DCT method	75x75	79.14%	74.94%	77.37%
	100x100	78.48%	77.37%	77.26%
	125x125	75.49%	75.48%	73.07%
	150x150	73.18%	74.83%	69.76%
	175x175	70.97%	74.50%	65.56%

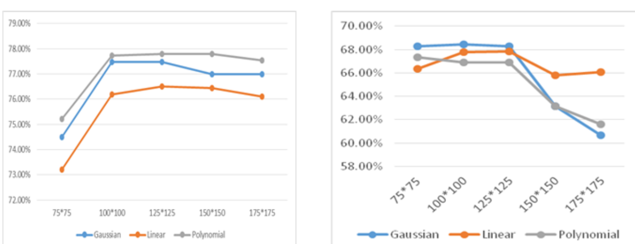


(a) The proposed OHATRS. (b) The benchmark ATRS.

Fig. 3. Classification Accuracies of (a) The Proposed OHATRS and (b) The Benchmark ATRS when Tested on Sub-Set (d) of the IFN/ENIT Database using Five Normalized Image Sizes and the SVM Classifiers.

TABLE. II. THE CLASSIFICATION ACCURACIES OF THE PROPOSED OFFLINE HATRS AND THE BENCHMARK ATRS WHEN TESTED ON SET (E) OF THE IFN/ENIT DATABASE USING THE SVM CLASSIFIERS

	Normalization size	Classification accuracy		
		Gaussian	Linear	Polynomial
Proposed system based on PCA	75x75	75.22 %	73.21%	74.50%
	100x100	77.72%	76.19%	77.47%
	125x125	77.80%	76.51%	77.47%
	150 x150	77.79%	76.43%	76.99%
	175x175	77.55%	76.11%	76.99%
Benchmark system using the DCT method	75x75	67.34%	66.37%	68.30%
	100x100	66.93%	67.81%	68.46%
	125x125	66.93%	67.82%	68.30%
	150x150	63.15%	65.81%	63.15%
	175x175	61.62%	66.05%	60.66%



(a) The proposed OHATRS. (b) The benchmark ATRS.

Fig. 4. Classification Accuracies of (a) The Proposed OHATRS and (b) The Benchmark ATRS when Tested on Sub-Set (e) of the IFN/ENIT Database using Five Normalized Image Sizes and the SVM Classifiers.

V. CONCLUSIONS AND FUTURE DIRECTIONS

This study proposed a holistic, multi-stage, system for recognition of the handwritten Arabic text that is based on PCA and SVM classifiers (OHATRS). Text recognition in this proposed system is performed at three stages: preliminary processing, feature extraction using PCA, and classification using the Gaussian, linear, and polynomial SVM classifiers. At the preliminary stage, text skeleton is extracted and text image is normalized into a uniform size for extraction of the universal Arabic word features using PCA. These extracted features are then used to classify the handwritten Arabic words using SVM classifiers.

The herein proposed OHATRS and the benchmark ARTS were evaluated on the (e) and (d) sub-sets of the IFN/ENIT database using (i) five word image normalization sizes (75x75, 100x100, 125x125, 150x150, and 175x175) and (ii) the Gaussian, linear, and polynomial SVM classifiers. The best classification accuracies (89.96% and 77.80%) produced by the proposed OHATRS were achieved using the 125x125 image normalization size and the Gaussian SVM classifier when this system was evaluated on the (d) and (e) sub-sets of the IFN/ENIT database, respectively. On the other hand, the best classification accuracies (79.14% and 68.46%) produced by the benchmark ARTS were achieved using the 75x75 and 100x100 image normalization sizes, respectively.

This study finds that the Arabic text recognition results of the proposed OHATRS are promising; the system produced better classification accuracies than the benchmark ATRS. As was highlighted in the foregoing section, the testing outcomes support effectiveness of the proposed system in recognition of the handwritten Arabic words. The experimental results underscore that the 125x125 word image size is the best size for optimum system performance and recognition results. For similar future studies, the researcher suggests training the proposed OHATRS using a combination of structural and statistical features.

REFERENCES

[1] A. Al-Shatnawi and K. Omar, "Methods of Arabic language baseline detection -the state of art". IJCSNS, 8 (10), 2008.

[2] A. Al-Shatnawi, "A Non-Iterative Thinning Method Based on Exploited Vertices of Voronoi Diagrams", PhD Thesis, University Kebangsaan Malaysia, Malaysia, 2010.

[3] A. Al-Shatnawi, "A Novel Baseline Estimation Method for Arabic Handwritten Text Based on Exploited Components of Voronoi Diagrams", International Arab Journal of Information Technology (IAJIT), 13(3), 2016.

[4] A. Al-Shatnawi, "A Preprocessing Model for Hand-Written Arabic Texts Based on Voronoi Diagrams", International Journal of Computer Science and Information Technology (IJCSIT), 7(6), (p.p 1-18) December 2015.

[5] A. Al-Shatnawi, "A skew detection and correction technique for Arabic script text-line based on subwords bounding", In Computational Intelligence and Computing Research (ICCIC), IEEE International Conference on, (pp. 1-5), IEEE, 2014.

[6] A. Al-Shatnawi, A. S. Safwan, F. AL-Zawaideh, and K. Omar, "Offline Arabic Text Recognition–An Overview". World of Computer Science and Information Technology Journal (WCSIT), 1(5), (pp.184-192), 2011.

[7] A. Al-Shatnawi, and K. Omar, "The Thinning Problem in Arabic Text Recognition-A Comprehensive Review", International Journal of Computer Applications, 103(3), (pp. 0975-8887), 2014.

[8] A. Da Rocha Gesualdi, and J. M. de Seixas, "Character recognition in car license plates based on principal components and neural processing", In VII Brazilian Symposium on Neural Networks Proceedings, (pp. 206-211), IEEE, (November, 2002).

[9] A. K. Hassan, and Alawi. M, "Proposed Handwriting Arabic Words classification Based on Discrete Wavelet Transform and Support Vector Machine", Iraqi Journal of Science, 58(2C), (pp.1159-1168), 2017.

[10] A. Tharwat, "Principal component analysis-a tutorial", International Journal of Applied Pattern Recognition, no. 3 (pp.197-240), 2016.

[11] B. El Qacimy, A. Hammouch, and M. A Kerroum, "A review of feature extraction techniques for handwritten Arabic text recognition", In 2015 International Conference on Electrical and Information Technologies (ICEIT), (pp. 241-245), IEEE. 2015.

[12] B. El Qacimy, M. A Kerroum, and A. Hammouch, “Word-based Arabic handwritten recognition using SVM classifier with a reject option”, in 2015 15th international conference on intelligent systems design and applications (ISDA), (pp. 64-68), IEEE, December 2015.

[13] F. Nashwan, M. A. Rashwan, H. M Al-Barhamtoshy, S. M. Abdou, and A. M. Moussa, ”A holistic technique for an Arabic OCR system”, Journal of Imaging, 4(1), 6, 2018.

[14] G. A. Abandah, K. S. Younis, and M. Z. Khedher, “Handwritten Arabic character recognition using multiple classifiers based on letter form”, In Proceedings of the 5th International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA) (pp. 128-133), February 2008.

[15] H. Nemmour, and Y. Chibani, ”Handwritten Arabic word recognition based on Ridgelet transform and support vector machines”, In 2011 International Conference on High Performance Computing & Simulation, (pp. 357-361), 2011.

[16] J. H. AlKhateeb, “Word based off-line handwritten Arabic classification and recognition: design of automatic recognition system for large vocabulary offline handwritten Arabic words using machine learning approaches”, Ph.D. dissertation, University of Bradford, 2010.

[17] K. Bansal, and S. Arora, “Face Recognition using PCA & LDA Algorithms,” Second International Conference on ACCT, (pp. 251-254), 2012.

[18] K. Jayech, M. A. Mahjoub, and N. E. Amara, “Synchronous multi-stream hidden markov model for offline Arabic handwriting recognition without explicit segmentation”, Neurocomputing, (pp. 958-971), 2016.

[19] K. Khan, R. Ullah, N. A Khan, and K. Naveed, “Urdu character recognition using principal component analysis”, International Journal of Computer Applications, 60(11), 2012.

[20] K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space", Philosophical Magazine, 2 (11), (pp.559–572), 1901.

[21] M, El-Bashir, “Principal components analysis for Arabic sub-word recognition”, 2010 International Conference on Intelligent Network and Computing (ICINC 2010) (pp. 432–434), 2010.

[22] M. Amara, K. Ghedira, K. Zidi, and S. Zidi, “A comparative study of multi-class support vector machine methods for Arabic characters recognition”, In 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA), (pp. 1-6), IEEE, (November 2015).

[23] M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, and H. Amiri, “IFN/ENIT-database of handwritten Arabic words”, In Proc. of CIFED, Vol. (2), (pp. 127-136), October 2002.

[24] M. S. Kadhm, and A. P. Abdul, “Handwriting word recognition based on SVM classifier”, International Journal of Advanced Computer Science & Applications, 1, (pp.64-68), 2015.

[25] M. S. Khorsheed, “Off-line Arabic character recognition–a review”, Pattern analysis & applications, 5(1), (pp. 31-45), 2002.

[26] M. T. Parvez, and S. A Mahmoud, “Offline Arabic handwritten text recognition: a survey”, ACM Computing Surveys (CSUR), 45(2), 23, 2013.

[27] M. W. Sagheer, C. L. He, N. Nobile, and C. Y. Suen, “Holistic Urdu handwritten word recognition using support vector machine”. In 2010 20th International Conference on Pattern Recognition, (pp. 1900-1903), IEEE, August 2010.

[28] O. B. Ali, and A. Shaout, “Hybrid Arabic Handwritten Character Recognition Using PCA and ANFIS”, In International Arab Conference on Information Technology (ACIT’2016), 2016.

[29] V. Märgner, and H. El Abed, “ICDAR 2009 Arabic handwriting recognition competition”, In 2009 10th International Conference on Document Analysis and Recognition, (pp. 1383-1387). IEEE, (July 2009).

[30] V. Vapnik, ”Statistical learning theory”, Vol. 1, 1998, New York: Wiley.