Artificial Immune Algorithm for Handwritten Arabic Word Recognition

Hassiba Nemmour and Youcef Chibani
Faculty of Electronic and Computer Sciences, University of Sciences and Technology
Houari Bouemediene, Algeria

Abstract: *In this work, a system for solving handwritten Arabic word recognition is proposed. The aim is focused on holistic word recognition, which is devoted to recognize averaged size lexicons by using a single classifier. Presently, we investigate the applicability of the Artificial Immune Recognition System (AIRS) to achieve the recognition task. For the feature generation step, ridgelet transform and pixel density features are combined to highlight both linear singularities and topological traits of Arabic words. Experiments are conducted on a vocabulary of twenty-four words extracted from the IFN/ENIT dataset. The results show that feature combination improves the recognition accuracy with more than 1%. The comparison with Support Vector Machine (SVM) classifier highlights the effectiveness of AIRS. This latter achieves comparable and sometimes better performance than SVM and can be extended to recognize any number of classes.*

1. Introduction

Word recognition is known as one of the most challenging tasks of handwriting analysis. In addition to be very sensitive to the writer properties, words are fairly complex patterns which suffer from the great variability in the writing style. Also, datasets are collected from real-life handwritten documents such as checks, forms, official contracts, mail envelopes, etc. For that reason, they are characterized by multi-script aspect as well as different writing conditions and tools. Specifically, there are two approaches for handwritten word recognition [22, 30]. The first approach, called analytical, employs the segmentation into sub-words or isolated characters. Then, words are recognized through character recognition. The analytical recognition is useful for problems with very large vocabulary where it becomes impossible to refer each word to a class of interest. However, it constructs a specific classifier for each word. The second approach is based on global or holistic analysis in which, words are considered as a single unit. In the present work, we are interested in the holistic approach, which is appropriate for problems with medium and small vocabulary such as address and literal amount recognition in bank checks [15, 22]. From a processing point of view, the main advantage of this approach is that it captures all-co articulation and variability effects into word images that are handled by the same classifier [35].

Recall that advances in handwritten word recognition are due to the use of robust artificial intelligence methods such as Hidden Markov Models (HMM), artificial neural networks and Support Vector

Machines (SVMs) [15, 22, 35]. Specifically, HMM are employed in the analytical approach for solving very large vocabulary problems. For medium vocabulary cases based on the holistic approach, all classifiers can be used. Precisely, research works reveal that SVMs provide the best performance [17, 19]. Several types of features were used to deal with the high variability within words. Note for instance, shape features such as contours, and edges; structural features such as ascenders, and loops, as well as statistical transforms such as Principal Component Analysis (PCA) and wavelets [10, 14, 16, 22]. Despite this, word recognition still remains a challenging recognition problem. In fact, it is very difficult to find optimal features since the complexity is related to the number of classes and the language specificities.

The present work is focused on Arabic language, which is substantially different from other languages. Specifically, Arabic script is semi-cursive because a word can include many cursive sub-words or connected-components [10]. In addition to dots, there are several diacritical marking such as dumma ('), hamza (ء), or madda (~). Moreover, a letter can take several shapes according to its position within the word [6, 30]. As example, the letter "Aïn", can be written through four shapes that are: "ع، ع، عـ، ـعـ". Also, some letters such as "djim: ج", "ha: ح" and "kha: خ", differ only in dots number and take exactly the same shape. Because of these characteristics, structural features that are computed on skeletons of connected components such as the number of loops and concavities constituted the conventional features for Arabic writing [15]. Roughly, research works in this field are based on methods that were already used for English and French

word recognition. Thereby, HMM were extensively used especially for problems with very large vocabulary [1, 6, 12, 15, 22, 29]. For average lexicon applications, a single classifier is used to perform the recognition task. Thereby, artificial neural networks and SVMs classifiers were extensively employed over the past years [10]. Nevertheless, the classifier accuracy is class-dependent, which lets the holistic word recognition an ongoing research issue. In order to obtain some improvement, various works employed classifier combination [15].

Despite this, classifier accuracies obtained for average number of words is significantly smaller than the human performance. Therefore, a robust recognition can be achieved by involving new classification schemes.

In this paper, we investigate the applicability of Artificial Immune Recognition System (AIRS) for solving handwritten Arabic word recognition. AIRS is a bio-inspired classification algorithm, which was introduced by Watkins *et al.* [36, 37]. First, applications of this algorithm were designed for solving binary pattern recognition problems such as thyroid diagnosis [23] and fault detection [24]. Then after, AIRS were extended for some multiclass problems such as classification of remotely sensed images [39]. Recently, some research works attempted to use the artificial immune system for solving handwriting recognition tasks. For instance, we note handwritten Russian uppercase letter recognition [38], English letter recognition [25] and handwritten character recognition [13]. Presently, AIRS is used for solving handwritten Arabic word recognition.

The remain of this paper is arranged as follows: Section 2 gives a brief review of the related works on Arabic word recognition. Section 3 describes the proposed recognition system while experimental results are presented and discussed in section 4. The last section gives the conclusion of this work.

2. Related Works

Researches on Arabic language started in the sixteen of the last century. However, Arabic word recognition systems emerged at the end of the eighteen. Specifically, Al Muallim and Yamagushi [5] proposed the first Arabic word recognizer. From this work, various systems have been proposed; but they were all tested on a private datasets, which prevents comparison between them. This changed in 2002 when the first volume of IFN/ENIT dataset was introduced [4]. Abuhaiba *et al.* [3] proposed the first Arabic word recognition system using this dataset. The system proceeds by segmentation into sub-words or connected components which are then, represented by direct straight-line approximation. Then, after a large number of recognition systems were developed using the different versions of this dataset. Most of them were

developed using HMM classifiers and various structural, statistical and topological features [2, 8, 9, 22, 29, 31, 32, 34]. An interesting overview on HMM-based Arabic word recognition was introduced by Srihari and Ball in [32]. In the holistic approach, that is adequate for medium sized number of classes, several classifiers can achieve segmentation-free recognition. Recall that such approach includes all methods that consider each word image as single entity upon, which features are extracted. Broumandnia *et al.* [10] investigated the use of wavelet packets and neural networks for solving Arabic word classification with 16 classes. On the other hand, Farah *et al.* [15] proposed a system for Arabic literal amount recognition in bank checks. Authors showed that it is necessary to proceed through classifier combination to enhance the results of K-Nearest Neighbors (K-NN) and the artificial neural network. Recently, in many pattern recognition applications with medium sized number of classes, SVMs provided higher performance than artificial neural networks and HMM [17, 19]. Besides, several works investigated the use of SVM for Arabic word recognition. Alalshekmubarak *et al.* [4], SVM was developed to recognize a lexicon of 24 classes using grid features. Furthermore, in [21] SVM were used with Discrete Cosine Transform (DCT) and PCA transforms to recognize a lexicon of 7 classes. Also, in [26] SVM were used with the Ridgelet transform. Unfortunately, recognition scores of these systems are altered by the complexity of Arabic language. Moreover, the holistic recognition approach is lexicon dependent since a single classifier should be developed to recognize several words. In this case, the recognition accuracy decreases when the lexicon size increases. Thereby, research efforts are devoted to develop more robust classifiers and feature characterization schemes. Presently, we investigate the applicability of AIRS for Arabic word recognition. The AIRS was introduced by Watkins [36] specifically for solving pattern recognition problems. In contrast to SVM, which needs a specific multi-class implementation, the AIRS can be used to solve any number of classes since it consists of developing a Memory Cells (MC) population that is used to classify data according to a K-NN rule. Moreover, compared to other learning machines, the AIRS is very simple to implement because data training is a one shot analytical process. Therefore, in this work AIRS are applied and investigated comparatively to solve handwritten Arabic word recognition for problems with medium lexicons.

3. Proposed System

This section describes the proposed Arabic word recognition system that is composed of feature generation and classification modules.

3.1. Feature Generation

For feature generation, statistical and topological features are used to characterize words. As reported in Figure 1, for statistical features the uniform grid with pixel densities is used. The grid is obtained by partitioning word images into a fixed number of regions (i.e., cells) with the same size [28]. Then, each cell is substituted by its pixel density, which yields a feature vector that contains N_R components (N_R: Is the number of cells in the word image). Statistical features are obtained using the Ridgelet transform that is used to highlight linear singularities within images. This is carried out by performing wavelet decomposition after the computation of radon transforms [11]. Presently, digital Ridgelet features are computed as follows:

1. Normalizing signature images to the same size (200×200 pixels).

2. Compute the approximate radon transform which is generally, described in general by:

$$RD(\theta, r) = \iint f(x, y)\delta(x\cos\theta + y\sin\theta - r)dxdy \qquad (1)$$

Where δ: Dirac distribution, θ: Angular variable, r: Radial variable.

The discrete radon transform for images is computed by using the FFT a follows:

- Compute the 2D FFT of the image.
- Cartesian to polar conversion.
- Compute the 1D inverse FFT on each angular line.

Ridgelet features are obtained by applying 1D Wavelet (1DW) transform along the radial variable in the radon space [11]. Figure 2 shows the radon as well as the Ridgelet transform for a handwritten Arabic word. The wavelet is computed for 3 decomposition levels, which yields the approximation coefficient and three details.
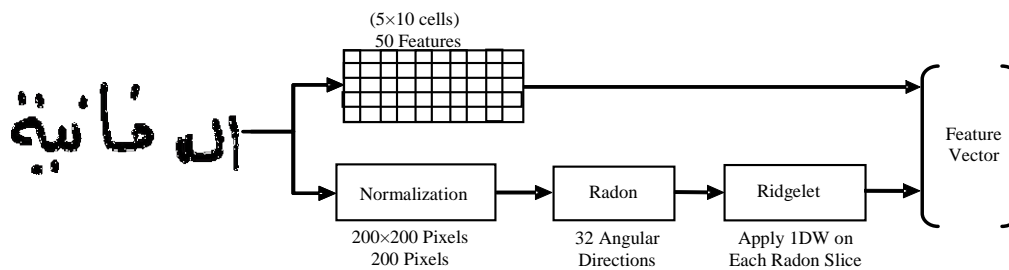

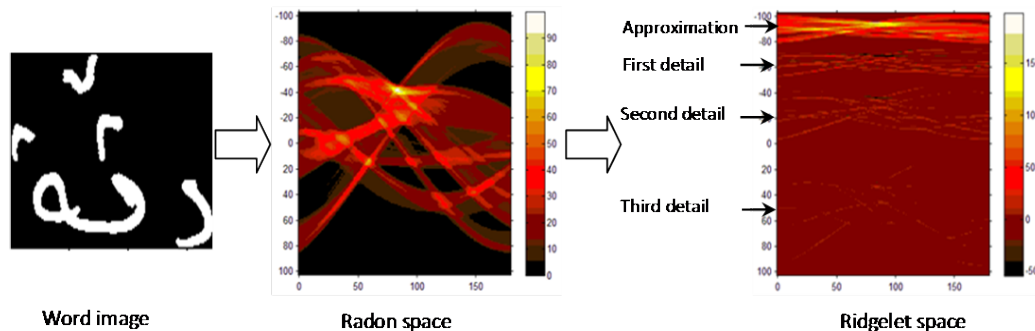Figure 1. Feature generation for Arabic words.

Figure 2. Illustration of ridgelet features for a training sample.

Recall that the Ridgelet transform has been successfully used in various image processing applications such as texture classification [7], face recognition [20], and printed Chinese character recognition [11]. Unfortunately, there are no references on the use of Ridgelet transform for handwriting recognition but similar transforms such as the Curvelet transform have been employed for handwritten Bangla character recognition [4]. Also, the Ridgelet transform was used in handwritten signature verification [27]. Presently, it is used for improving handwritten Arabic word recognition with medium lexicon.

3.2. Artificial Immune Recognition System

Artificial Immune Systems (AIS) are inspired from the natural immune system that employs B cells in the production and secretion of antibodies, which are specific proteins that bind to the antigen [33]. As other bio-inspired methods such as artificial neural networks and genetic algorithms, AIS emerged in 1990s. However, they started to gain a real success in pattern classification field in the last recent years. This is due to the AIRS, which adapted the AIS to multiclass pattern recognition and classification [36].

Roughly, in the AIRS (and most AIS techniques) the idea of antigen/antibody binding is employed and is known as antigenic presentation [33]. When dealing with learning algorithms, this is used to implement the idea of matching between training data (antigens) and potential solutions (B-Cells). Once the affinity between a B-Cell and an antigen has been determined, the B-Cell involved transforms into a plasma cell and experiences clonal expansion. During the process of

clonal expansion, the B-Cell undergoes rapid proliferation (cloning) in proportion to how well it matches the antigen [36]. The clonal expansion and affinity maturation are employed to encourage the generation of potential MC. These MC are later used for classification. Specifically, the AIRS training is composed of the following steps.

1. *Initialization*: In this step data are normalized such that Euclidian distance scales in the range [0, 1]. Then, an Affinity Threshold (AT) is computed according to the following equation:

$$AT = \frac{\sum_{i=1}^{n} \sum_{j=i+1}^{n} affinity\left(ag_i, ag_j\right)}{n(n-1)/2} \qquad (2)$$

Where ag_i: ith antigen which is the feature vector of a training sample, *affinity*: Euclidian distance, n: number of antigens (training data).

The initialization step is achieved by seeding of initial MC and initial ARB population. In other words, for each class of interest, some training samples are randomly selected to be used as prototypes of their classes. These selected samples are called antibodies or established MC.

1. *Training process*: The training of each datum (antigen: ag_i) is performed as follows:

- *MC-Match Selection*: The MC-Match is the MC, which has the highest stimulation to this antigen. The stimulation "ST" is defined as:

$$ST\left(ag_i, MC\right) = 1 - ED\left(ag_i, MC\right) \qquad (3)$$

Where *ED*: Euclidian distance.

- *Clonal expansion*: In this step, MC-Match is used to generate a set of randomly mutated clones.
- *Resources competition*: Generated clones undergo a competition based on their stimulations with respect to the current training sample. According to the number of allowed resources by the algorithm, clones which have small stimulations are removed.
- *MC-Candidate selection*: After resources competition, among the remaining clones, the one which has the highest stimulation to the current antigen will be selected as a potential memory cell. This clone is called MC-Candidate.
- *MC set introduction*: in this step, MC-Candidate is compared to MC-Match based on their stimulations with respect to the current antigen. If MC-Match has a highest stimulation, the MC-Candidate is unknown and the training process jumps to the next antigen. Otherwise, the stimulation between MC-Candidate and MC-Match is computed. If they are too much similar MC-Candidate replaces MC-Match in the MC

population. In the opposite case, MC-Candidate is added while keeping MC-Match.

3. *Classification step*: Once the training stage is finished, test samples are classified according to a K-NN decision computed with respect to the MC population. For further details on AIRS, reader can be referred to [35, 36].

4. Experimental Analysis

The proposed recognition system was evaluated on samples extracted from the well-known IFN/ENIT database [18]. This dataset contains 26400 images of Tunisian town names. Words are written by more than 411 scripts using different writing tools. To construct a medium sized vocabulary, we collected 24 words corresponding to 24 classes of interest grouping names with the largest appearance frequencies. The selected dataset has several particularities, which make its training a challenging task. As shown in Table 1, some words differ only in one letter. For instance, classes 9 and 10 differ in the fourth letter as well as classes 11 and 12 that differ in the first letter). In addition, some samples of classes 7 and 14 are composed of a word and digits while those of the classes 13, 15 and 16 are composed of two independent words. Besides, data contain high variability in size and in the intra-word space.

Table 1. Selected word classes.

Class 1	Class 2	Class 3
انا ضور	شتعاك	لش ماة
Class 4	**Class 5**	**Class 6**
الخليج	الرضاع	ا ك ودة
Class 7	**Class 8**	**Class 9**
المنزه و	هارت	الفكت
Class 10	**Class 11**	**Class 12**
الفجة	نقة	د ت ـ
Class 13	**Class 14**	**Class 15**
تل الغزلان	المارزه ١٥	راس الذراع
Class 16	**Class 17**	**Class 18**
سيدى الظاهر	ذ تو ثى	الفحص
Class 19	**Class 20**	**Class 21**
الفـايض	اله خانية	الشوامخ
Class 22	**Class 23**	**Class 24**
خزبص	الشرايع	خـا ل

Presently, the experimental design of the proposed Arabic word recognizer is two folds. The first consists of finding the optimal parameter selection for Ridgelet features. The second step concerns parameter optimization and performance assessment for the AIRS algorithm.

4.1. Feature Optimization

Parameter selection for data features was investigated using SVM classifiers. The SVM implementation is based on the one-against-all approach with the radial basis function kernel. SVM parameters (The kernel standard deviation and the regularization parameter) were selected based on cross-validation upon training samples. Recall that the Ridgelet transform is obtained by applying the 1DW decomposition on the discrete radon features. Thereby, it employs several setup parameters, which require an experimental tuning. We note in particular, the radon matrix size and the wavelet parameters. So, in first test, the influence of radon matrix size, which is controlled by the number of angular directions θ and radial projections r was evaluated.

Roughly, θ varies in the range [0°, 180°] while r is designated by the size of images (it corresponds to the longest axis within the word image). Therefore, in order to obtain the same size for all data, images were normalized into square matrix of 200×200 pixels. Recognition rate variations according to the size of radon matrix are plotted in Figure 3. It is easy to see that the matrix of 32×32 components makes the best trade-off between recognition accuracy (83.5%) and the size of the Ridgelet vector which has in this case, 1024 components. It is worth noting that a large radon matrix yields a very large Ridgelet feature vector and the corresponding recognition rate is weak.
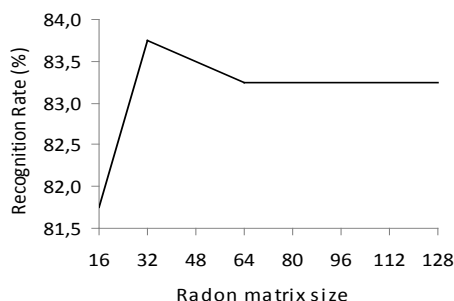

Figure 3. Ridgelet recognition rates obtained for various sizes of the radon matrix.

Once radon's parameters selected, the next step to obtain Ridgelet coefficients is the wavelet decomposition. The wavelet requires several setup choices such as the wavelet function, the number of decomposition levels and the selection of approximation or details coefficients. Therefore, several tests were performed to find the optimal selection. Here, for space consideration, we present the most significant tests. Note that, after each decomposition level, we obtain an approximation and a detail coefficient. Besides, the number of possible decomposition levels depends on the initial size of radon matrix. In fact, for a matrix of 32×32 components the number of decomposition levels can go to 5. Nevertheless, the recognition rate assessment showed that after the 3rd decomposition level of

"Daubechies 1" wavelet, the Ridgelet transform keeps approximately the same performance. This supplied 1024 Ridgelet features among, which there are 512 approximation features. Furthermore, experimental results showed that when switching between approximation and details, the recognition rate varies significantly. As shown in Figure 4, if we consider only the approximation, the recognition accuracy is altered to 70% while details allow a recognition rate of 83.5%. This means that details allow the same recognition rate as the whole wavelet vector (approximation and all details). Thereby, details coefficients were selected to generate word features since they make the best tradeoff between the recognition accuracy and the feature size, which has 512 components.
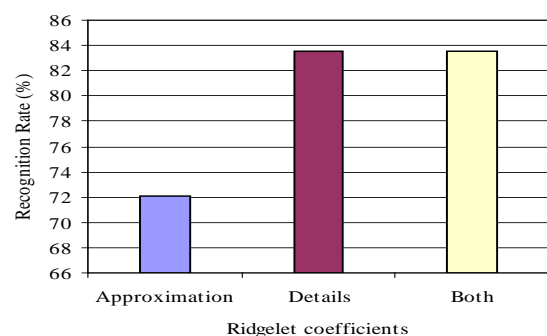

Figure 4. Recognition rate for different selections of the Ridgelet coefficients.

Finally, the feature vector of handwritten words is composed of 50 grid features that are concatenated with 512 detail coefficients of the Ridgelet transform. Then, the remaining tests were devoted to optimize parameter selection for the AIRS algorithm.

4.2. AIRS for Arabic Word Recognition

The adopted AIRS has several setup parameters, which should be experimentally tuned. Note that there is no theoretical rule to be followed when selecting these parameters. For that reason, several run-passes were performed to find the best value for each of them. Specifically, in the AIRS implementation, the following parameters should be experimentally tuned:

- *Stimulation Threshold*: Used as stopping criteria for the training process. It scales in the range [0, 1].
- *Resources Number*: An integer that expresses the total importance of all mutated clones.
- *Hyper mutation Rate and Mutation Rate*: Used to express the mutation probability of each component in the feature vector.
- *Affinity Threshold*: Corresponds to the global similarity between all training data. It is used in the comparison between MC-Candidate and MC-Match.
- *Clonal Rate*: An integer which controls the number of generated clones.

Experiments performed by varying these parameters showed that only the clonal rate, resources number as well as the K value in the K-NN decision have a significant effect on the recognition accuracy. Thereby, in what follows the AIRS behavior is analyzed according to these parameters.

The clonal rate influence was investigated by considering values between 1 through 300. From Figure 5 we can see the compromise between the runtime and the recognition performance. Specifically, small clonal rates produce few cloned samples, which are not sufficient to describe the variability within word classes. This leads to low recognition scores with a fast training duration. On the contrary, larger values induce a large training time and better recognition accuracy. More precisely, the best recognition rate that is about 85% is achieved with a clonal rate that equals 150.
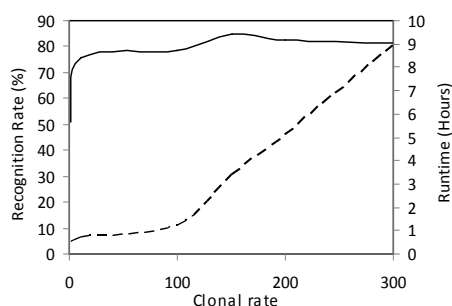

Figure 5. AIRS performance according to the clonal rate variations.

Furthermore, Figure 6-a lots variations of the recognition rate according to resources number. It reveals that up to 150, the recognition rate increases from 72.3% to 84.8%. Nevertheless, higher values (larger than 150) alter the AIRS performance. It is worth noting that the increase of this parameter means that we keep more mutated clones among which the MC-Candidate is selected. Thereby, after resources competition, it is important to keep sufficiently large population to allow an adequate selection of MC-Candidate. Furthermore, Figure 6-b shows that the recognition accuracy is substantially altered when the K-NN rule takes more than one neighbor.

a) According to resources number.

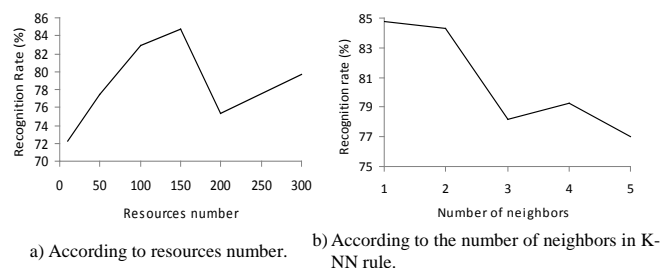b) According to the number of neighbors in K-NN rule.

Figure 6. Recognition rate variations.

From 1 to 5 neighbors, about 7% are lost in the recognition accuracy. This outcome can be explained by the fact that the intra-class variability is very high where each new memory cell describes a specific part of this variability. Consequently, each test sample exhibits high similarity to a specific memory cell of a given class but its average similarity to all MC of this class is much smaller.

Finally, the performance assessment of AIRS was carried out using the parameter selection that is reported in Table 2. The AIRS was compared to SVM based on recognition rates and runtime. For time assessment, we consider the training duration and the recognition speed that is defined by the number of recognized samples per second. For SVM, the couple (σ=5, C=10) was obtained through cross-validation over the training set. Table 3 summarizes results obtained for both SVM and AIRS. The recognition rate was evaluated by considering grid and Ridgelet features individually and then by considering their combination.

Table 2. Parameter selection for AIRS.

AIRS parameter	Value
Stimulation Threshold	0.92
Resources Number	150
Mutation Rate	0.01
Hypermutation rate	0.1
Affinity Threshold Scalar	0.5
Clonal Rate	150
Number of Neighbors in K-NN	1

Table 3. Performance evaluation for handwritten Arabic word recognition.

SVMs			
	Zoning	Ridgelet	Both
Recognition Rate (%)	83.1	83.9	84.3
Training Time (Hours)	0.01	0.2	0.25
Recognition Speed	125.2	46.3	67.8
AIRS			
	Zoning	Ridgelet	Both
Recognition Rate (%)	84.5	79.8	85,2
Training Time (Hours)	3.5	7.6	9.8
Recognition Speed	252	56.4	71.2

As can be seen, the SVM classifier provides almost similar performance with both topological (grid features) and statistical features (Ridgelet transform). For this reason, when considering the combination, the recognition improvement is about 0.5%. On the contrary, the AIRS achieve its best accuracy with grid features. However, with Ridgelet transform the recognition accuracy is less than 80%. This outcome can be explained by the lack of suitability between Euclidian distance that is used in affinity and stimulation computation and Ridgelet values which belong in the range [-1, 1]. This issue was circumvented by the combination with grid features. Thereby, the AIRS outperform SVM with approximately 1% in the overall recognition rate. Furthermore, as expected the runtime evaluation reveals that SVM is much faster than AIRS. In fact, the training time of AIRS grows with the clonal rate parameter as well as with the size of data which make the mutated clones generation a time consuming task. Nevertheless, this is not considered as drawback since both classifiers have approximately the same recognition speed, which is the most important factor.

We remark that both classifiers have approximately the same recognition speed with all features. This outcome reveals that in real time recognition the AIRS can be as fast as SVM.

5. Conclusions

In this paper, we proposed a system for handwritten Arabic word recognition that is devoted for solving medium vocabulary recognition problems. In this system, the applicability of the AIRS was investigated while data features were obtained by combining topological grid features with the Ridgelet transform. Experimental protocol was composed of words extracted from the well-known IFN-ENIT dataset. The performance evaluation of AIRS was conducted comparatively to the SVM classifier. From all experiments, we conclude that AIRS is suitable for solving medium vocabulary Arabic word recognition. This classifier outperforms SVM in terms of recognition accuracy but it requires a larger training time. Also, it has been shown that despite of using several set-up parameters, only the clonal rate and the resources number require a careful selection to achieve a satisfactory performance. On the other hand, the results obtained for grid and Ridgelet features reveal that AIRS is more adapted with data that scale in the range [0, 1]. This behavior reflects limitations of the Euclidian distance, which is used in affinity and stimulation measures. Consequently, higher precision could be obtained when substituting the Euclidian distance by more powerful metric. Another idea that could be fruitful to bring more robustness into AIRS decision consists of replacing the K-NN rule by a trained decision function.

References

[1] Al-Badr B. and Mahmoud S., "Survey and Bibliography of Arabic Optical Text Recognition," *Signal Processing Journal*, vol. 41, no. 1, pp. 49-77, 1995.

[2] Al-Maàdeed S., Higgens C., and Elliman D., "Recognition of Off-Line Handwritten Arabic Words Using Hidden Markov Model Approach," *in Proceeding of the16th International Conference on Pattern Recognition*, Quebec, pp. 481-484, 2002.

[3] Abuhaiba I., Mahmoud S., and Green R., "Recognition of Handwritten Cursive Arabic Characters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 664-672, 1994.

[4] Alalshekmubarak A., Hussein A., and Wang Q., "Off-Line Handwritten Arabic Word Recognition Using SVMs with Normalized Poly Kernel," *in Proceeding of the International Conference on*

Neural Informtion processing*, Berlin, pp. 85-91, 2012.

[5] Al Muallim H. and Yamaguchi S., "A Method of Recognition of Arabic Cursive Handwriting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 715-722, 1987.

[6] Amin A., "Off-line Arabic Character Recognition: the State of the Art," *Pattern Recognition Letters*, vol. 31, no. 5, pp. 517-530, 1998.

[7] Arivazhagan S., Ganesan L., and Kumar T., "Texture Classification Using Ridgelet Transform," *Pattern Recognition Letters*, vol. 27, no. 16, pp. 1875-1883, 2006.

[8] Assma O., Khalifa O., and Hassan A., "Handwritten Arabic Word Recognition: A Review of Common Approaches," *in Proceeding of the International Conference on Computer and Communication Engineering*, Kuala Lumpur, pp. 801-805, 2008.

[9] Benouareth A., Ennaji A., and Sellami M., "Arabic Handwritten Word Recognition Using HMM with Explicit State Duration," *Eurasip journal on Advances in Signal Processing*, vol. 2008, pp. 1-13, 2008.

[10] Broumandnia A., Shanbehzadeh J., and Varnoosfaderani M., "Persian/Arabic Handwritten Word Recognition Using M-Band Packet Wavelet Transform," *Image Vision and Computing Journal*, vol. 26, no. 6, pp. 829-842, 2008.

[11] Chen G., Bui T., and Krzyzak A., "Rotation Invariant Feature Extraction Using Ridgelet and Fourier Transforms," *Journal of Pattern Analysis and Application*, vol. 9, no. 1, pp. 83-93, 2006.

[12] Chen M., Kundu A., and Zhou J., "Off-Line Handwritten Word Recognition Using A Hidden Markov Model Type Stochastic Network," *IEEE Transactions on Pattern analysis and Machine Intelligence*, vol. 16, no. 5, pp. 481-496, 1994.

[13] Chen Y., Liang C., Yang D., Peng L., and Zhong X., "A Handwritten Character Recognition Based on Artificial Immune Algorithm," *in Proceeding of the International Conference on Computer Application and System Modeling*, Taiyuan, pp. 273-276, 2010

[14] Dzuba G., Filatov A., Gershuny D., and Kill I., "Handwritten Word Recognition-the Approach Proved by Practice," *in Proceeding of the 6th International Workshop on Frontiers in Handwriting Recognition*, Taejon, pp. 99-111, 1998.

[15] Farah N., Souici L., and Sellami M., "Classifiers Combination and Syntax Analysis for Arabic Literal Amount Recognition," *Engineering Applications of Artificial intelligence*, vol. 19, no. 1, pp. 29-39, 2005.

[16] Guillevic D. and Suen C., "Cursive Script Recognition Applied to the Processing of Bank Cheques," *in Proceeding of the 3rd International Conference on Document Analysis and Recognition*, Montreal, pp.11-14, 1995.

[17] Hsu C. and Lin C., "A Comparison of Methods for Multi-Class Support Vector Machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415-425, 2002.

[18] IFN/ENIT-database Arabic OCR handwritten Arabic word recognition, www.ifnenit.com, Last Visited 2014.

[19] Justino E., Bortolozzi F., and Sabourin R., "A Comparison of SVM and HMM Classifiers in the off-line Signature Verification," *Pattern Recognition Letters*, vol. 26, no. 9, pp. 1377-1385, 2005.

[20] Kautkar S., Koche R., Keskar T., Pande A., Rane M., and Atkinson G., "Face Recognition Based on Ridgelet Transform," *Procedia Computer Science*, vol. 2, pp. 35-43, 2010.

[21] Khalifa M. and BingRu Y., "A Novel Word Based Arabic Handwritten System Using SVM Classifier," *Advanced Research on Electronic Commerce, Web Application and Communication in Computer and Information Science*, vol. 143, pp. 163-171, 2011.

[22] Knerr S., Augustin E., Baret O., and Price D., "Hidden Markov Model Based Word Recognition and its Application to Legal Amount Reading on French Checks," *Journal of Computer Vision and Image Understanding*, vol. 70, no. 3, pp. 404-419, 1998.

[23] Kodaz H., Ozsen, S., Arslan, A., and Gunes, S., "Medical Application of Information Gain Based Artificial Immune Recognition System Diagnosis of Thyroid Disease," *Journal of Expert Systems with Applications*, vol. 36, no. 2, pp. 3086-3092, 2009.

[24] Laurentys C., Palhares R., and Caminhas W., "A Novel Artificial Immune System for Fault Behavior Detection," *Journal of Expert Systems with Applications*, vol. 38, no. 6, pp. 6957-6966, 2011.

[25] Liang C., Peng L., Hong Y., and Wang J., "An English letter Recognition Algorithm Based Artificial Immune," *An English Letter Recognition Algorithm Based Artificial Immune*, vol. 5553, pp. 371-379, 2009.

[26] Nemmour H. and Chibani Y., "Artificial Immune System for Handwritten Arabic Word Recognition," *in Proceeding of the 3rd International Conference on Innovative Computing Technology,* London, pp. 463-466, 2013.

[27] Nemmour H. and Chibani Y., "Off-Line Signature Verification Using Artificial Immune Recognition System," *in Proceeding of the 10th International Conference on Electronics Computer and Computation*, Ankara, pp.168-171, 2013.

[28] Nemmour H. and Chibani Y., "Training Tangent Similarities for improving Ridgelet Characterization of Handwritten Arabic Words," *in Proceeding of the 11th International Conference on Intelligent Systems Design and Applications*, Cordoba, pp. 1050-1055, 2011.

[29] Pechwitz M., Al Abed H., and Märgner V., *Guide to OCR for Arabic Script*, Springer, 2012.

[30] Plötz T. and Fink G., "Markov Models for Offline Handwriting Recognition : A Survey," *International Journal of Document Analysis and Recognition*, vol.12, no. 4, pp. 269-298, 2009.

[31] Sari T. and Sellami M., "Overview of Some Algorithms of Off-Line Arabic Handwriting Segmentation," *The International Arab Journal of Information Technology*, vol. 4, no. 4, pp. 289-300, 2007.

[32] Srihari S. and Ball G., *Guide to OCR for Arabic Script*, Springer, 2012.

[33] Timmis J. and Neal M., "A Resource Limited Artificial Immune System for Data Analysis," *Journal of Knowledge-based Systems*, vol. 14, no. 3, pp. 121-130, 2001.

[34] Touj S., Amara N., and Amiri H., "Arabic Handwritten Words Recognition Based on a Planar Hidden Markov Model," *The International Arab Journal of Information Technology*, vol. 2, no. 4, pp. 318-325, 2005.

[35] Vinciarelli A., "A Survey on Off-Line Cursive Word Recognition," *Journal of the Pattern Recognition Society*, vol. 35, no. 7, pp. 1433-1446, 2002.

[36] Watkins A., "AIRS:A Resource Limited Artificial Immune Classifier," *Master Thesis on Computer Sciences*, Mississippi state university, 2001.

[37] Watkins A., Timmis J., and Bogess L., "Artificial Immune Recognition System (AIRS): An Immune-Inspired Supervised Learning Algorithm," *Genetic Programming and Evolvable Machines*, Kluwer Academic Publishers, vol. 5, no. 3, pp. 291-317, 2004.

[38] Yang Y., "Application of Artificial Immune System In Handwritten Russian Uppercase Character Recognition," *in Proceedings of the International Conference on Computer Science and Service System*, Nanjing, pp. 238-241, 2011.

[39] Zhang L., Zhong Y., and Li P., "Applications of Artificial Immune Systems in Remote Sensing Image Classification," *in Proceeding of Congress of International society for Photogrammetry and Remote Sensing*, Istanbul, pp. 397-401, 2004.

Hassiba Nemmour received the M.S. and PHD degrees from the Faculty of Electronic and Computer Sciences, at USTHB University, Algiers, in 2003 and 2010. She is currently a senior lecturer at the same faculty. Her research interests include artificial intelligence, machine learning and soft computing applied to the handwriting recognition field.

Youcef Chibani received his Magister and Doctoral in electrical engineering from the University of Science and Technology Houari Boumedienne, Algiers, Algeria. Currently, he is working as a teacher-researcher, as a full professor since 2007. His research interest includes feature extraction, multidirectional decomposition, image enhancement, neural network, support vector machines, one-class classifiers, combination of multiple classifiers and its application for document analysis, character recognition, character segmentation, writer identification, signature verification, historical document restoration and word spotting. He was a head of the Scientific Council and responsible for master and doctoral training. He co-authored many papers in international peer-reviewed journal and conferences. Since 2015, he is director of the Communicating and Intelligent System Engineering Laboratory.