

Use an Efficient Neural Network to Improve the Arabic Handwriting Recognition

Husam Ahmed Al Hamad

Department of Information Technology
College of Computer, Qassim University
Qassim, Saudi Arabia
hhamad@qu.edu.sa, hushamad@yahoo.com

Abstract—Using an efficient neural network for recognition and segmentation will definitely improve the performance and accuracy of the results; in addition to reduce the efforts and costs. This paper investigates and compares between results of four different artificial neural network models. The same algorithm has been applied for all with applying two major techniques, first, neural-segmentation technique, second, apply a new fusion equation. The neural techniques calculate the confidence values for each Prospective Segmentation Points (PSP) using the proposed classifiers in order to recognize the better model, this will enhance the overall recognition results of the handwritten scripts. The fusion equation evaluates each PSP by obtaining a fused value from three neural confidence values. CPU times and accuracies are also reported. Experiments that were performed of classifiers will be compared with each other and with the literature.

Keywords: Arabic recognition, segmentation, neural network.

I. INTRODUCTION

The concept of handwriting has existed for old time, for the purpose of expanding people's memory and facilitating communication together, much of culture may be attributed to the advent of handwriting, many researchers began to focus their attention on attempting to simulate intelligent behavior, one such example was the attempt to imitate the human ability to read and recognize printed and handwritten characters [1]. The term "handwriting" is defined as meaning of artificial graphic marks contains some message through the mark's relation to language [2].

Pattern recognition that is study of how machines can observe the environment, learn to distinguish patterns of interest from their background, and make sound and reasonable decisions about the categories of the patterns, the best pattern recognizers in most instances are humans, yet we do not understand how humans recognize patterns [3]. The aim of this research is allow a comparison between four different Artificial Neural Networks (ANNs), then discover which is better for recognition of the off-line Arabic handwritten script.

II. HISTORICAL BACKGROUND

ANNs History starts from produced Hebbian that learning with a mechanism of neural plasticity in 1940. Then researchers develop the first linear classifier of training that

called perceptron, it is an essentially classifier. After that and in 1960 a multi-layered model was derived. At first, the use of the Multi-Layer Perceptron (MLP) was complicated by the lack of a suitable learning algorithm [4]. In 1975 Kunihiko Fukushima [5] has designed multilayered neural network with a training algorithm. Network structure and the methods that used are interconnection weights change from one neural to another. Networks can propagate information in one direction only. In 1986 Rumelhart [6] the application area network of back-propagation algorithm are gained recognition and utilized multiple layers of weight-sum units of the type $f = g(w \cdot x + b)$. Training was done by a form of stochastic gradient descent. In 1982 Kohonen [7] introduced the Self-Organizing Map (SOM) network model. SOM is a special kind of technique which organizes itself according to the input patterns that it is trained with. As known the SOM originated from the LVQ (Learning Vector Quantization) network that introduced as an idea by Kohonen's in 1972. In 1988 Broomhead and Lowe [8] were introduced the Radial Basis Function (RBF) networks. Although, this network was developed thirty years ago with another name is potential function method, the work by Broomhead and Lowe opened a new frontier in the neural network community.

III. OFF-LINE VS. ON-LINE RECOGNITION

Off-line handwriting recognition refers to the process of recognizing words that have been scanned from a paper or book and are stored digitally format. After being stored, it is conventional to perform further processing to allow superior recognition. The main approaches that exist for off-line Arabic handwriting recognition may be divided into segmentation based and holistic ones. In general, the former approach uses a strategy based on the recognition of individual characters or patterns whereas non-segmentation based deals with the recognition of the word image as a whole [9]. In the on-line case, the handwriting is captured and stored in digital form via different means. Usually a special pen is used in conjunction with an electronic surface. As the pen moves across the surface or paper, the two-dimensional coordinates of successive points are represented as a function of time and are stored in order [2].

It is generally accepted that the on-line method of recognizing handwriting has achieved better results than off-line until now, this may be attributed to the fact that more

information may be captured in the on-line case such as the direction, speed and the order of strokes of the handwriting, this information is not easy to recover from handwritten words written on a non-digital medium such as paper [1].

IV. PROBLEMS OF ARABIC SCRIPTS

As is known, Arabic is spoken by millions of peoples and important in the culture of many more, and many research efforts have been published in the area of recognition of Arabic handwritten script, until now these efforts have not reached good results due to the following reasons: (1) Arabic words are overlapped and written always cursively, i.e., more than one character can be written connected to each other. (2) Arabic uses many types of external objects, such as “dots”, “Hamza”, “Madda”, and diacritic objects. These reasons make the task of line separation and segmentation scripts more difficult. (3) Arabic characters can have more than one shape according to their position: initial, middle, final, or standalone [10].

V. ARTIFICIAL NEURAL NETWORKS

ANNs are a novel approach that follows a different ways from traditional computing methods to solve problems; they are much more powerful because they can solve problems that we do not exactly know how to solve, they are a recent development tool that is tries to simulate some properties of biological neural network. In other words, ANNs have the ability to learn complex nonlinear input/output relationships. Neural network composed of several layers each layer is composed of a large number of interconnected elements called neurons working in unison to solve specific problems [3], To be used as classifiers, Neural Networks need to be trained on a set of data whereby the training algorithm modifies the weights between each connection until some criteria is met. The different usage of artificial neural network area are including speech recognition, intrusion detection systems, image analysis and adaptive, virus detection, pattern recognition (face, image, script, fingerprint, noise...), robot control, in order to construct software agents in computer and video games or autonomous robots.

A. *Types of Neural Networks (ANN)*

Many models of artificial neural network had been proposed to model the learning mechanism of brain; the widely applied models will be described in this paper. The most common family of neural networks for pattern classification recognition is Feed-Forward Back-Propagation network (FFBP) it is very simple and effective to implement, it was first established by Rumelhart [6], it has been applied successfully to different application domains, such as pattern recognition, controlling, prediction, system identification, etc. [11], the weight inputs transmits to the neurons in the first layer and the neurons transmits their outputs to the neurons in to the next layer, etc., the network not contain any cycles or loop as an advantage [12]. Another popular type of network is Multilayer Perceptron (MLP), in many applications such as filtering, noise removal, pattern recognition, it has been widely used coupled with the back-propagation (BP) algorithm [13] and it is highly nonlinear in its parameters; the neurons are organized in the form of layers and the weights connecting the

neurons in successive layers, therefore, it requires a training procedure calculated based on the training samples and target classes [14]. Radial-Basis Function (RBF) is also the most often used of networks type; a RBF is a feed-forward neural network that has only one hidden layer with an unsupervised training method [15], it founds to be very attractive for many computing problems and used in many different research fields, such as function approximation, regularization, noisy interpolation and pattern recognition [16, 17], it has a very compact topology and its learning speed is very fast and easy because of its locally tuned neurons [18]. Finally, one of the widely applied networks is also the Self-Organizing Map (SOM) network [19], the abstract relation between the input signal and the synaptic adaptation of neurons was mathematically described by Kohonen in 1982 [20, 21], the input layer can have different dimensions and topology, it learns from high dimensional data and maps them on a low, data close together in high-dimensional space will be close in the mapped low-dimensional space, such abilities make the SOM being widely applied in data visualization and clustering tasks. [22, 23].

VI. RESEARCH TECHNIQUES

The aim of this research as mentioned before is allow a comparisons between four different models of neural networks in order to validate the prospective segment points (PSP) which were added by Arabic Heuristic Segmentor (AHS) technique as a primitives points in the word image. Recognition of characters that will be used in the phase of neural-segmentation is also applied, this technique outputs confidence values consider as an input for fusion phase to decides which segment point is valid or invalid, according this determining which network will consider the better. The technique uses an Arabic heuristic segment to over-segment the handwriting based on extracting the shape feature of the word image, in other words; AHS adds an exceed segment points more than word need. Following this, a neural confidence-based module is used to evaluate a prospective segmentation point by obtaining a fused value from three neural confidence values: segment point validation (SPV), right character validation (RCV), and central character validation (CCV) [10]. Therefore, the technique uses the recognition to validate each prospective segment point. The algorithms were built used Matlab v.2010, and the computer was used has a core 2due, 2.40GHz processor, 4GB memory and uses windows 7-32bit operating system.

A. *Handwriting Database*

The database that was used for majority of segmentation experiments in this research were obtained from twenty different persons. For training networks only 620 characters were extracted by the first ten persons, and for testing networks exactly 500 random words were extracted from two paragraphs contained all possibility shapes of Arabic characters that wrote by the rest persons. In this research, all words images were converted to binary format, more details are seen in [10].

B. *Arabic Heuristic Segmentor (AHS)*

AHS is a new feature-based heuristic segmentor technique that proposed by Al Hamad [10]. The AHS technique partitions

the handwritten words into primitives (over-segmentation) that will then be processed further to provide the best segmentation, in other words, the technique employs various heuristics to decide how to split the word into segments, thereafter, three specific errors are calculated: *over segmented*, *missed*, and *bad* segmentation point.

Table 1 illustrates number of segmentation errors that were obtained for characters only. It may be noted that for the results listed below, an *over-segmentation* was recorded if the body of the character was divided into more than three segments; however, ligature segments surrounding the character were not taken into account, on other words, over-segmentation refers to a character that has been divided into more than three components. The *missed* refers to the eventuality that two touching characters have not been separated at all, missed error occurs when no segmentation point is found between two successive characters. Finally, *bad* refers to segmentations that are neither correct nor missed, it occurs if for example two touching characters have been split in such a way that either one or both of the characters have been disfigured or particular character components have been incorrectly separated, this means bad error refers to a segmentation point that could not be used to extract a character precisely.

TABLE 1. OVER-SEGMENTATION PERFORMANCE OF AHS WITH 3349 SEGMENTATION POINTS SP.

SP	Correct Segmentation	Over-Segmentation Error Rates			
		Over-segmented	Missed	Bad	Total
3349	2832	49	9	459	517
	84.56%	1.46%	0.27%	13.71%	15.44%

C. Direction Feature Extraction

A novel feature extraction technique is used for extracting the structural features of handwritten characters. Direction feature extraction technique combines local feature vector and global structural information and provides integrated features to a neural network for training and testing. The proposed approach first employs an existing character outline tracing technique, which traces the contour of a given character image. Then, the directions of line segments comprising the characters that are detected and the foreground pixels are replaced with appropriate direction values. Finally, features of the characters based on the location of background to foreground pixel transitions are extracted and neural training and classification is performed.

D. Neural Networks

Each network needs numbers of vectors for training, and each vector array extract from one character as mentioned before. In this research two neural networks are used, the first network is trained with features extracted from segment area (SA) originally located by the heuristic algorithm, the network verifies whether each particular area is or isn't characteristic of a segment point, the second network is trained with direction features that are extracted from right character (RC) and central character (CC) of prospective segmentation point. In many ways, this has been seen as an advantage i.e. faster and more accurate training. Neural networks have been thoroughly

training with 620 patterns, and tested on 500 words [10] for each of the four proposed models: Feed-Forward Back-Propagation (FFBP), Multilayer Perceptron (MLP), Radial-Basis Function (RBF), and finally Self-Organizing Map (SOM) network.

E. Fusion Confidence Values

Fusion confidence value is a set of equations that calculate the final results of each segment point, it decides whether the segment point is valid or invalid based on confidence value of the neural network, two possible types of fusion equations are employed, first, calculate the Correct Segmentation Point (CSP) where Segmentation Point Validation (SPV) ≥ 0.5 as shown in equation (1), second, calculate the Incorrect Segmentation Point (ISP) where $SPV < 0.5$ as shown in equation (2), finally, the outcome of fusion decision is decided based on determining the maximum value between the CSP and ISP as shown in equation (3), these equations indicate a procedure for computing the final confidence value for each segment point in the word image, the greater confidence out of both decides the final identity of the SP being examined. If the CSP confidence is greater, the segmentation point will be set as being correct. Conversely, if the ISP confidence prevails as being larger, the SP will be discarded and no longer used in further processing. The entire technique involves an analysis of each word from right to left.

$$f_{CSP}(ft1, ft2, ft3) = f_{SPV_Ver}(ft1) + f_{RCC_Ver}(ft2) + (1-f_{CCC_Ver}(ft3)) \quad (1)$$

$$f_{ISP}(ft1, ft2, ft3) = (1-f_{SPV_Ver}(ft1)) + f_{RCC_Ver}(ft2) + f_{CCC_Ver}(ft3) \quad (2)$$

$$f(\text{confidence}) = \max [(CSP), (ISP)] \quad (3)$$

Where, $f_{SPV_Ver}(\text{features})$ is confidence value of SPV that it is an output of neural network, $f_{RCC_Ver}(\text{features})$ is a confidence value for right character, and $f_{CCC_Ver}(\text{features})$ is confidence value for center character (reject neuron output).

VII. EXPERIMENTAL RESULTS

The experiment obtains the results by verifying the prospective segmentation points according to heuristic segmenter technique that was used to over-segment the handwriting. Two results are obtained, following sub-sections show them.

A. Results of the Character Recognition

All set of experiments results of recognition for the four types of neural networks that implemented in this research are displayed in Table 2. The table contains details about the training errors of the networks, CPU time and the classification rate on the test set. Numbers of epochs that were trained for all networks were 300 epochs, the table illustrates results of characters recognition for FFFB, MLP, RBF, and SOM networks, with 620 training/testing pairs using direction feature with 120 inputs for each character/area, and the CPU time of training characters.

TABLE 2. EXPERIMENT RESULTS OF CHARACTERS RECOGNITION WITH 620 TRAINING/TESTING PAIRS USING DIRECTION FEATURE AND 120 INPUTS.

Neural Network	Training Error	CPU time (Second)	Classification	
			Accuracy	Rate Set
FFBP	10.48%	56.1448	78.06%	484/620
MLP	1.45%	449.4233	72.58%	450/620
RBF	1.13%	103.5379	95.32%	591/620
SOM	13.93%	202.3333	24.35%	151/620

The above results describe the recognition rate and CPU time for all types of neural networks that trained and tested by the same database and same feature extraction. On the other hand, to enhance the classification rates, the number of testing set must be increase at least two-fold or three-fold; this number help to improve the accuracy of overall segmentation, Figure 1 illustrates the characters recognition rates of all different networks.

B. Results of the Neural-based Segmentation Technique

After training and testing the neural networks, and calculate recognition rate, next step is validate all prospective segment points. The results of the neural-based segmentation technique are calculated based on number of correct and incorrect identified SP in word samples. Neural network verifies whether segmentation points are valid or invalid based on neural confidence-based module for validating the prospective segmentation points. If network output a height confidence value this indicates a point is a valid segmentation point; a low confidence value indicates a point should be ignored. The accuracy of the AHS, networks, and fusion equations can impact on the efficiency and accuracy of the overall segmentation techniques. Table 3 shows the results of the neural-based segmentation technique for all models.

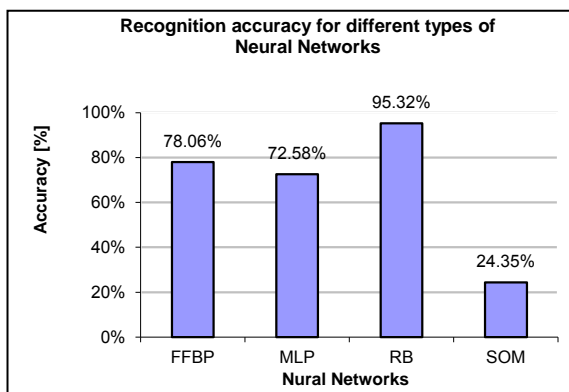


Fig 1. Characters recognition rates for all different neural networks.

Figure 2 shows the performance of all networks, include correctly and incorrectly of validate / invalidate the segment points. Results of correctly and incorrectly identified of valid and invalid segmentation points need more study and analysis to clearly explain the reasons of recognition and segmentation rates, and explain the behavior of each network. The final experiments presented here are used direction feature to locate segmentation points from word image based on the neural

network algorithm. These experiments give more accurate view of the execution of the segmentation technique, as its deals with the same input patterns for training. Figure 3 illustrates the successfully and non-successfully segmentation of word images using heuristic segmentation with and without segmentation point validation.

TABLE 3. PERFORMANCE RESULTS OF THE NEURAL-BASED SEGMENTATION TECHNIQUE.

Neural Network	Result	Correctly Identified		Incorrectly Identified	
		Valid	Invalid	Valid (Bad)	Invalid (Missed)
FFBP	Count	2233	250	743	123
	%	66.68%	7.46%	22.19%	3.67%
	Total	2483		866	
	%	74.14%		25.86%	
MLP	Count	459	2350	89	858
	%	13.71%	70.17%	2.66%	25.62%
	Total	2439		910	
	%	72.83%		27.17%	
RBF	Count	1567	767	263	752
	%	46.79%	22.90%	7.85%	22.45%
	Total	2334		1015	
	%	69.69%		30.31%	
SOM	Count	2336	33	969	11
	%	69.75%	0.99%	28.93%	0.33%
	Total	2369		980	
	%	70.74%		29.26%	

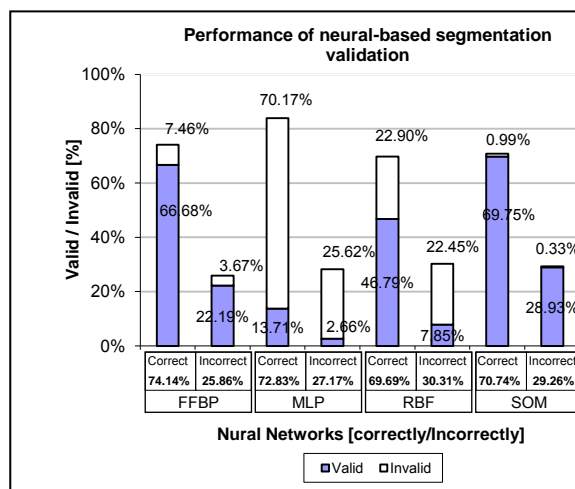


Fig 2. Correctly/Incorrectly identify valid/invalid of the neural-based segmentation.

Results of correctly and incorrectly identified of valid and invalid segmentation points need more study and analysis to clearly explain the reasons of recognition and segmentation rates, and explain the behavior of each network.

The final experiments presented here are used direction feature to locate segmentation points from word image based on the neural network algorithm. These experiments give more accurate view of the execution of the segmentation technique, as its deals with the same input patterns for training. Figure 3 illustrates the successfully and non-successfully segmentation

of word images using heuristic segmentation with and without segmentation point validation.

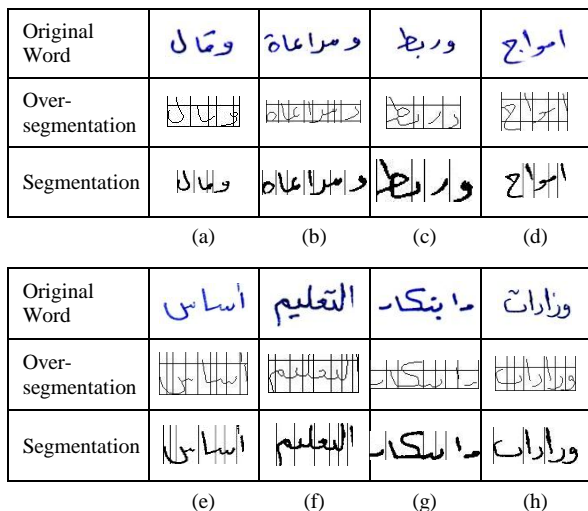


Fig 3. Sample of word images segmented, (a-d) successful segmentation, and (e-h) unsuccessful segmentation.

VIII. ANALYSIS AND COMPARISON OF RESULTS

The purpose of this section is to analyze the experimental results that obtained in this research as well discuss their significance. As we have seen from table 2, there are a clearly different between the recognition rate, training error and CPU time between all different networks. RBF network has a best rate of recognition with approximate double time of FFBP, which also has a good recognition rate; MLP also reached to accepted recognition rate, but recognition rate of SOM is worst one. The observed results for overall segmentation that have seen in table 3 are not fairly.

After analysis the recognition and segmentation results, what is seen at the first outlook a difference or discrepancy between the results of recognition and segmentation, to analyze the variation we need to review fusion equations that used, in addition, not be overlooked that the final segmentation step is to determine which segment points are correctly kept (valid) or removed (invalid), and which segment points are incorrectly kept (valid) or removed (invalid). In other words, verify the segment points that have been added in a phase of AHS. Further analysis of the results will contain a comprehensive study that describes the impact all parts of the technique, such as networks that used, training set, testing set, fission equations. On the other hand, the outputs that obtained as final results from segmentation phase may not be easily compared with other researchers in the literature, as in most cases segmentation accuracy for Arabic handwritten words is either not measured at all, or is measured with respect to the number of correct or incorrect segmentations found. This means there are many researchers have not tested the techniques of segmentation point's verification or over-segmented Arabic words. Table 4 contains a summary of the literature results that compared with the results of this paper.

TABLE 4. COMPARISON OF THE SEGMENTATION RESULTS FROM THE LITERATURE.

Accuracy	Language	Year	Data set	Reference
85.70%	Cursive English handwriting	1995	50 real mail envelopes	K. Han [24]
90.00%	Printed English handwriting	1996	Alphanumeric characters	S-W. Lee [25]
75.90%	Cursive English handwriting	1997	CEDAR database	B. Eastwood [26]
81.21%	Cursive English handwriting	1997	Griffith University database	Blumenstein and Verma [27]
75.28%	Cursive English handwriting	2000	CEDAR database	Myer Blumenstein [1]
86.90%	Cursive English handwriting	2000	CEDAR database	G. Nicchiotti [28]
69.72%	Arabic handwriting	2001	Local database (360 addresses or 4000 words)	Alaa Hamid [Error! Bookmark not defined.]
85.74%	Cursive English handwriting (Testing 1031 from 1718 SP)	2005	CEDAR database	Chun Ki Cheng [29]
85.00%	Arabic handwriting (Sub-words segmentation)	2008	Local database (200 images)	Jawad AlKhateeb et al. [30]
82.98%	Arabic handwriting	2010	Local database (500 words)	Husam Al Hamad [10]

CONCLUSION

The results of character recognition of Radial-Basis Function network is reached to 95.32%, the result is very favorably with minimal training error and less CPU time, and high performance. Upon comparison, it is produces an acceptable recognition accuracy. The integral part in this research is validation of segmentation point, it is developed to remove invalid segmentation points and retain the valid point in order to increase accuracy of overall segmentation. The final validation results of segmentation points using direction feature and new equations of fusion are 74.14% for FFBP network, and 69.69% for RBF. The accuracy of neural-base segmentation technique was unfavorable. Fusion equations that have been used need more study and development to obtain better results such as the high results that have been obtained by recognition phase.

REFERENCES

[1] Myer Blumenstein, "Intelligent Techniques for Handwriting Recognition", School of Information Technology, PhD Dissertation, Griffith University-Gold Coast Campus, Australia, 2000.

- [2] R. Plamondon, S. N. Srihari, "On-line and Off-line Handwriting Recognition: A Comprehensive Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 63–84, 2000.
- [3] Jayanta Kumar Basu, Debnath Bhattacharyya, Tai-hoon Kim, "Use of Artificial Neural Network in Pattern Recognition", *International Journal of Software Engineering and Its Applications*, vol. 4, No. 2, pp. 24-32, 2010.
- [4] Rosenblatt, F., "The Perceptron: A Probabilistic Model For Information Storage and Organization In The Brain". *Psychological Review*, vol. 65, pp. 386-408, 1958.
- [5] Fukushima, Kunihiko, "Cognitron: A self-organizing multilayered neural network", *Biological Cybernetics* vol. 20(3-4), pp. 121–136, 1975.
- [6] Rumelhart, David E., Hinton, Geoffrey E., Williams, Ronald J., "Learning representations by back-propagating errors", *Nature*, vol. 323(6088), pp. 533–536, 1986 .
- [7] T. Kohonen, "Self-organized formation of topologically correct feature maps", *Biological Cybernetics*", vol. 43(1), pp. 59-69, 1982.
- [8] D.S. Broomhead, D. Lowe, "Multivariate functional interpolation and adaptive networks", *Complex Systems*, vol. 2, pp. 321-355, 1988.
- [9] Xiaolong Fan, Brijesh Verma, "Segmentation vs. Non-Segmentation Based Neural Techniques for Cursive Word Recognition: An Experimental Analysis", *International Journal of Computational Intelligence and Applications*, vol. 2(4), pp. 377–384, 2002.
- [10] Husam A. Al Hamad, Raed Abu Zitar, "Development of an Efficient Neural-based Segmentation Technique for Arabic Handwriting Recognition", *Pattern Recognition journal - ELSEVIER*, vol. 8(43), pp. 2773-2798, 2010.
- [11] J. Bilski, "The Ud Rls Algorithm for Training Feedforward Neural Networks", *Int. 1. Appl. Math. Comput. Sci.*, pp. 115-123, 2005.
- [12] Abdalla, O.A., Zakaria, M.N., Sulaiman, S., Ahmad, W.F.W, "A comparison of feed-forward back-propagation and radial basis artificial neural networks: A Monte Carlo study", *Information Technology (ITSim)*, vol.2, no., pp.994-998, 2010.
- [13] S. A Durai, E., A Saro. "Image compression with back-propagation neural network using cumulative distribution function", *World Academy of Science, Engineering and Technology*, vol. 17, pp. 60-64,2006.
- [14] Ebeid, H.M., "Using MLP and RBF neural networks for face recognition: An insightful comparative case study", *International Conference on Computer Engineering & Systems (ICCES)*, vol., no., pp.123-128, 2011.
- [15] S. Chen, C. F. N. Cowan, and P. M Grant, "Orthogonal least squares learning algorithm for radial basis function networks", *IEEE Trans. Neural Networks*, vol. 2, no. 2, pp. 302-30,1991.
- [16] AS. Khalifa, R.A Ammar, M. F. Tolba, T. Fergany, "Dynamic online allocation of independent task onto heterogeneous computing systems to maximize load balancing", *8th IEEE International Symposium on Signal Processing and Information Technology*, art. no. 4775659, pp. 418-425, 2008.
- [17] Zhang, Yong, Xue, ZhiMao, "RBF Neural Network Application to Face Recognition", *International Conference on Challenges in Environmental Science and Computer Engineering (CESCE)*, vol.2, no., pp.381-384, 2010.
- [18] R. V. Babu, S. Suresh, A Makur, "Robust object tracking with radial basis function networks", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. I, pp. 937-940, 2007.
- [19] T. Kohonen, "Self-Organizing Maps. Berlin: Springer-Verlag", 1997.
- [20] Kohonen, T., "The self-organizing map", *Proceedings of the IEEE*, vol.78, no.9, pp.1464-1480, 1990.
- [21] Deotale, N., Vaikole, S.L., Sawarkar, S.D., "Face recognition using artificial neural networks", *The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, , vol.2, no., pp.446-450, 2010.
- [22] M. PalHi, T. Honkela, T. Kohonen, "Bibliography of self-organizing map (SOM)", *papers: 2002-2005 addendum*, *Information and Computer*.
- [23] Hsin-Chang Yang, Chung-Hong Lee, "A novel self-organizing map algorithm for text mining", *International Conference on System Science and Engineering (ICSSE)*, vol., no., pp.417-420, 2010.
- [24] K. Han, I. K. Sethi, "Off-line Cursive Handwriting Segmentation", *Proceedings of the 3rd International Conference on Documents Analysis and Recognition*, pp. 894–897, 1995.
- [25] S-W. Lee, D-J. Lee, H-S. Park, "A New Methodology for Gray-Scale Character Segmentation and Recognition", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, pp. 1045–1051, 1996.
- [26] B. Eastwood, A. Jennings, A. Harvey, "A Feature Based Neural Network Segmentor for Handwritten Words", *International Conference on Computational Intelligence and Multimedia Applications, ICCIMA, Gold Coast, Australia*, vol. 97, pp. 286–290, 1997.
- [27] M. Blumenstein, B. Verma, "A Segmentation Algorithm used in Conjunction with Artificial Neural Networks for the Recognition of Real-World Postal Addresses", *International Conference on Computational Intelligence and Multimedia Applications, ICCIMA, Gold Coast*, vol. 97 , pp. 155–160, Australia, 1997.
- [28] G. Nicchiotti, C. Scagliola, "A Simple and Effective Cursive Word Segmentation Method", *Proceedings of the 7th International Workshop on Frontiers in Handwriting Recognition, Amsterdam*, pp. 499–504, 2000.
- [29] Chun Ki Cheng, Michael Blumenstein, "The Neural-based Segmentation of Cursive Words using Enhanced Heuristics", *ICDAR, 8th International Conference on Document Analysis and Recognition*, vol. 5, pp. 650–654, 2005.
- [30] Jawad H AlKhateeb, Jianmin Jiang, Jinchang Ren, Stan S Ipson, "Component-based Segmentation of Words from Handwritten Arabic Text", *Proceedings of World Academy of Science, Engineering and Technology, ISSN*, vol. 31, pp. 1307–6884, 2008.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en_US