The International Journal of Engineering and Information Technology

journal homepage:www.ijeit.misuratau.edu.ly

Offline Recognition System for Arabic Handwritten Words Using Artificial Neural Networks

M.M. Makhlou
Department of Networks and Communications, College of Information Technology, Misurata University
Misurata, Libya
m.makhlouf@it.misuratau.edu.ly

A.M. Abdulshahed
Department of Electrical and Electronic Engineering, College of Engineering, Misurata University
Misurata, Libya
a.abdulshahed@eng.misuratau.edu.ly

Abstract— **Arabic handwritten recognition is playing an important role in communicating and documenting information in everyday life. Handwriting has functional importance considering its ubiquity in human activities, as in reading handwritten notes in a PDA (Personal Digital Assistant), in postal addresses on envelopes, in amounts in bank accounts, etc. In this work, an offline recognition system for Arabic handwritten names is proposed using Artificial Neural Network (ANN). As part of the feature extraction, region props were used, it extracts array with fields: Area, Centroid and Bounding Box. A feed-forward neural network with three different topologies was trained using backpropagation algorithm. The number of nodes in the hidden layer was varied; the lower the number of hidden layers, the higher the recognition rate. Moreover, we found that the first topology (with two hidden layers) produced a higher recognition rate 92% from the trained data compared to the second and third topologies that produced a smaller recognition rate.**

Index Terms: **Offline Handwritten Recognition; Arabic handwritten; Artificial Neural Networks; Backpropagation Neural Network.**

I. INTRODUCTION

With the advent of emerging technology, handwriting has tended to persist as a way of communicating and documenting information in everyday life. Computer detection of handwriting has functional importance considering its ubiquity in human activities, as in reading handwritten notes in a PDA, in postal addresses on envelopes, in amounts in bank accounts, in handwritten fields of types, etc. [1]. Due to its numerous characteristics, the Arabic language is highly complicated. It has 28 letters and written in cursive from right to left, with the exception of six letters. The majority of the letters are linked to the one before them, the letter form complicates the script and its shape is determined by whether the letter is at the beginning, middle, or end [2]. Arabic handwritten names recognition can be classified into two parts which is online and offline. Both online and offline system are consisting of five main stages which is data acquisition, preprocessing, feature extraction, comparison process and decision. This research focuses on the offline system which could be considered as more challenging compare to the online system. This is due to the offline system did not capture the dynamic which can help the classifier as one of the most significant features. Hence, it will produce higher accuracy as stated in [3], [4]. In the offline system, only static image features are visible and can be considered.

Many researches have been conducted to develop Arabic handwritten recognition system. Elzobi et.al [5] proposes an un-constrained recognition approach for the handwritten Arabic script. Their approach starts by explicitly segmenting each word image into its constituent letters, then extracting feature vectors corresponding to different scales and orientations in the segmented image using a Gabor wavelet filter bank. Siddh et.al [6] presented combination statistical and structural approaches for character recognition. Firstly, the statistical method recognizes the main body of a character using modified direction features and Support Vector Machines. Secondly, in structural classification, dot descriptors are used to recognize the exact shape of an Arabic character. Typical handwritten recognition system is shown in Figure 1. Various groups have used a wide variety of common features. Until features can be extracted, many of these features involve normalization of the image.

In this field, other techniques have been developed, such as the Nearest Neighbor process, which was used on a small dataset containing 20 characters, but needed to resize the image. A technique called the radial method used to acquire statistical character features was defined by Peter Burrow. This technique may have very bad effects [7].

Received 30 Oct , 2021; revised 16 Nov, 2021; accepted 16 Jan, 2021.

Available online 28 Nov, 2021.

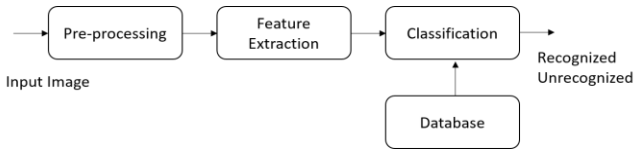

Figure 1. Typical Handwritten Recognition System.

On the classifier part, two methods have been mostly utilized, i.e., SVM and neural network [8] [9] [10]. And more recently, a deep learning neural network has been utilized in [10]. Although many researches have been conducted on offline Arabic handwritten recognition, but there are still many aspects have not been considered. In this work, we have collected our own Arabic handwritten name image database. Therefore, the objective of this paper is to develop an offline system of Arabic handwritten name using artificial neural network (ANN) to cope with the variation in the Arabic handwritten name images. The pre- processed and segmented handwritten image was passed to the feature extraction process. The extraction of features transfers the two-dimensional image into a vector set such that the input image is represented by a set of numerical values that are to be transferred to recognizer for recognition.

In this research, we use the whole pixel of the handwritten image, as image features and feed to the artificial neural network. Performance, recognition rate and training time of the proposed ANN were recorded and analyzed.

II. PROPOSED ARABIC HANDWRITTEN NAME RECOGNITION SYSTEM

This proposed algorithm, in which the system has two main parts, i.e., feature extraction and classifier. The feature extraction part is including the image segmentation, aligning and cropping and color to grayscale conversion. The classifier part is including the training and testing of the model with the developed image database (see Figure 2).

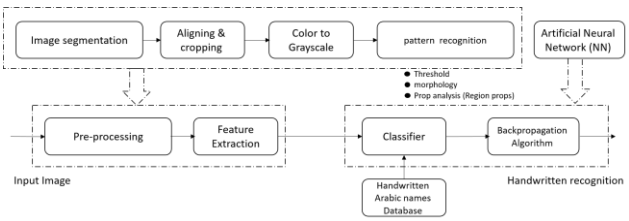

Figure 2. Proposed Offline Arabic Handwritten Name Recognition System.

A. Bwlabel and Regionprops for Extract Features in Matlab

In this paper, we used a MATLAB command "Bwlabel" (Label connected components in 2-D binary image) to mark the two-dimensional, two-value image of the connected group. Moreover, the MATLAB command "Regionprops" is applied to Measure properties of image regions.

L = Bwlabel (BW) returns the label matrix L that contains labels for the 8-connected objects found in BW. (8-connected) Pixels are connected if their edges or corners touch. Two adjoining pixels are part of the same object if they are both on and are connected along the horizontal, vertical, or diagonal direction.

B. Artificial Neural Network Classifier

The Artificial Neural Network (ANN) is a computational model based on the structure of the human brain, with functions and procedures based on it. By simulating the human brain, the artificial neural network (ANN) was developed as a learning tool. In order to decrease the time, it takes to solve a problem; a neural network is built up of numerous layers that should all be finite. The use of parallel functioning of neurons in ANN allows it to solve complex handwriting recognition problems Two of the most extensively utilized ANN designs for pattern recognition systems (RBFs) are Multilayer Perceptions (MLPs) and Radial Basis Functions (RBFs) [11] [12]. The number of hidden layer and nodes could be varied. Figure 4 shows the structure of feed-forward ANN that used in this research. While Figure 5 shows the MATLAB implementation.

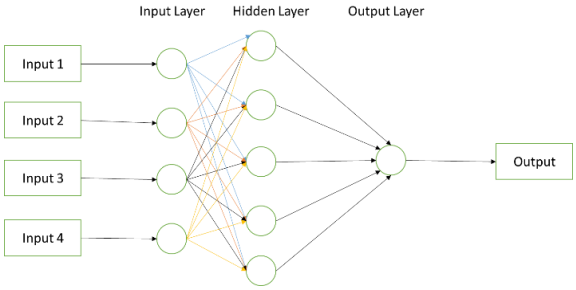

Figure 3. Feedforward Neural Network

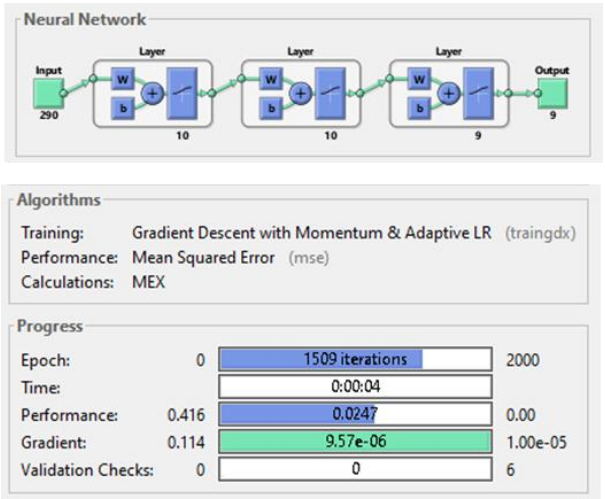

Figure 4. ANN MATLAB Implementation.

III. RESULTS AND DISCUSSION

In this section, the experimental setup, Arabic handwritten name image database, different features, different number of hidden layers and different number nodes in the hidden layer will be discussed.

A. *Experimental Setup*

A high-performance system was used for processing, i.e., a multicore system with Intel Core i7 4610M 3.00 GHz (4 cores with 8 threads), 8 G Bytes RAM, 500 G Bytes HDD hard disk, installed on Windows 10 operating system and MATLAB 2016b with Image Processing, Signal Processing and Neural Network Toolboxes.

In this experiment, a Neural Network toolbox in MATLAB has been used in order to train the model with pre-registered names. The input to the model is a word 'name' (not separated characters), then, the name converted to a vector format in order to find a perfect match with other pre-registered names. Three neural network topologies were configured with different hidden layers, and then trained to test them in order to recognize handwritten Arabic names. The input to the model is an image file and the final result saved on a "notepad" file format.

B. *Arabic Handwritten Name Image Database*

The Arabic Handwritten names are collected from two persons with 50 names has been taken from them with various styles. The database contains sample names, and each sample contains the name in four different patterns.

The Arabic handwritten names are taken using a painter software to convert them into digital format. The images that are taken will be cropped according to a guideline box. Then, the Arabic handwritten names that are collected will be divided into two folders as follows: 70% for training stage and 30% for testing stage. Figure 6 shows an example of collected Arabic handwritten names.

Figure 5. Sample of Arabic Handwritten Names.

C. *Performance Evaluation*

Table 2, illustrate that as the number of hidden layer increases, the recognition rate decreases. Furthermore, it also shows that about 92% recognition rates can be achieved when the hidden layers were only two hidden layers.

However, it can be noticed that a higher the number of hidden layers gives a lower recognition rate. Consequently, the number of hidden layers will affect the performance and recognition rate of the system. There are many differences in the image name database due to the different writing style. However, it was found that the first topology was the most suitable topology because it gives a high recognition rate. Table I shows each topology with its number of hidden layers and nodes. While Table II shows the effect of changing in performance and different on the training time and recognition rate in each topology.

Table 1. The topology's and it's number of hidden layers and node

Topology	**Input layer**	**Hidden layer**	**Number of nodes**	**Output layer**
Topology 1	one	two hidden layers	first hidden layer with 10 nodes and second with nodes equal to number of names in database= 50	one
Topology 2	one	three hidden layers	first hidden layer with 10 nodes, second hidden layer with 10 nodes and third with nodes equal to number of names in database= 50	one
Topology 3	one	four hidden layers	first hidden layer with 10 nodes, second hidden layer with 10 nodes, third hidden layer with 10 nodes and fourth with nodes equal to number of names in database= 50	one

Table 2. The effects of changing in performance and different on the training time and recohnition rate

Topology	**Dictionary**	**Performance (%)**	**Training Time (seconds)**	**Recognition Rate (%)**
Topology 1	20 names	1.03e-05	01	100 %
	25 names	9.81e-06	01	100 %
	30 names	0.000302	01	100 %
	35 names	0.000649	01	100 %
	40 names	0.0196	01	80 %
	45 names	0.000883	01	100 %
	50 names	0.0196	01	70 %
Topology 2	20 names	0.00864	02	60 %
	25 names	0.00323	01	90 %
	30 names	0.00474	03	90 %
	35 names	0.0102	01	30 %
	40 names	0.00785	04	60 %
	45 names	0.0102	04	40 %
	50 names	0.00586	04	90 %
Topology 3	20 names	4.63e-05	02	100 %
	25 names	0.000606	04	60 %
	30 names	0.0289	02	0 %
	35 names	0.00529	05	0 %
	40 names	0.0243	02	0 %
	45 names	0.0215	02	0 %
	50 names	0.0389	03	0 %

IV. CONCLUSIONS AND FUTURE WORKS

This work has presented an offline Arabic handwritten names recognition system using neural network. Approximately 50 name images from two different writing styles were collected. Three different topologies with various number of hidden layers have been constructed. A feed-forward neural network was used as classifier, and then trained using backpropagation algorithm for each topology. The number of nodes has been varied. Results indicate that the higher number of hidden layers shows lower recognition rate. Moreover, the training time was also higher. As a future work, this proposed system can be applied in useful real-life applications, such as recognize words and numbers in a bank check, and to convert a handwritten text into an editable file format.

ACKNOWLEDGMENT

First and foremost, I give my deep thanks to Allah for giving me the opportunity and the strength to accomplish this work.

I would like to thank my father and mother, my brothers and sisters, my husband and his family who showered me with their support and showed me unconditional love. To my supervisor Dr. Ali Abdulshahed for his help and support during my work for creating a research environment that has been very inspiring. This work would not have been possible without his guidance, encouragement and motivation. He deserves my respect and thanks.

All those who participated directly or indirectly to the realization of this work.

REFERENCES

[1] Mars, A. and G. Antoniadis (2016). "*Arabic online handwriting recognition using neural network.*" International Journal of Artificial Intelligence and Applications (IJAIA) 7(5).

[2] Butt, H., M. R. Raza, M. J. Ramzan, M. J. Ali and M. Haris (2021). "*Attention-Based CNN-RNN Arabic Text Recognition from Natural Scene Images*." Forecasting 3(3): 520-540.

[3] Perwej, Y. (2013). "*Recurrent Neural Network Method in Arabic Words Recognition System.*" arXiv preprint arXiv:1301.4662.

[4] Alginahi, Y. (2017). "*Document Processing and Arabic Optical Character Recognition: A User Perspective*." International Journal of Computer Science and Information Security 15.

[5] Elzobi, M., A. Al-Hamadi, Z. Al Aghbari, L. Dings and A. Saeed (2014). Gabor wavelet recognition approach for off-line handwritten arabic using explicit segmentation. Image Processing and Communications Challenges 5, Springer: 245-254.

[6] Siddhu, M. K., M. T. Parvez and S. N. Yaakob (2019). Combining statistical and structural approaches for Arabic handwriting recognition. 2019 International conference on computer and information sciences (ICCIS), IEEE.

[7] Burrow, P. (2004). "*Arabic handwriting recognition.*" Report of Master of Science School of Informatics, University of Edinburgh.

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

[9] MAMOUNI, E. M. (2017). Recognition of Printed Arabic Script Using Support Vector Machine, Ahmed Draia University-Adrar.

[10] M. Shams, A. Elsonbaty, and W. ElSawy, "*Arabic Handwritten Character Recognition based on Convolution Neural Networks and Support Vector Machine*," arXiv preprint arXiv:2009.13450, 2020.

[11] N. Altwaijry and I. Al-Turaiki, "*Arabic handwriting recognition system using convolutional neural network*," Neural Computing and Applications, vol. 33, pp. 2249-2261, 2021.

[12] K. RamaKrishna, C. V. K. Rao, V. A. Ramam, and R. S. Rao, "*Mathematical Neural Network (MaNN) Models Part V: Radial basis function (RBF) neural networks (NNs) in Chemometrics, Envirometrics and Medicinometrics (ChEM)*," Journal of Applicable Chemistry, vol. 3, pp. 1807-1893, 2014.