ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF RESEARCH IN PHARMACEUTICAL SCIENCES

Published by JK Welfare & Pharmascope Foundation Journal Home Page: www.ijrps.com

Using Artificial Intelligence System for Pattern Identification Based on Features Extracted from Gel Electrophoresis Images

Sarmad M. Hadi*[1], Al-Faiz M Z[2], Ali A. Ibrahim[1]

[1]Al-Nahrain university-college of information engineering-information and communication engineering department- Iraq-Baghdad
[2]Al-Farahidy University-Iraq-Baghdad

Article History:

Received on: 06.09.2019
Revised on: 17.12.2019
Accepted on: 24.12.2019

Keywords:

neural networks,
linear vector quantization,
DNA,
gel electrophoresis,
Back-Propagation

ABSTRACT

Artificial intelligence has many branches of image processing-based applications in terms of classification and identification, error back-propagation neural network is a great match for such applications as long as linear vector quantization (LVQ) and pattern recognition is another great match for recognition of digital images based on their features. The dataset used in this paper are gel electrophoresis images where 6 features had been extracted from the images and used as input to a neural network for learning and then checked for recognition purposed and the system managed to recognize all the 6 images. Six features had been used: average, standard deviation, smoothness, skewness, uniformity, and entropy. A tiny error rate where allowed in the recognition program to cover the variation of the dataset and the test data (gel-electrophoresis images). The proposed system had successfully managed to identify all of the learned data in both LVQ and error-back-propagation. Error-back-propagation proved itself as a great tool in terms of learning time compared with LVQ, which was very slow in terms of learning time and recognition.

*Corresponding Author

Name: Sarmad M. Hadi
Phone: +9647506896479
Email: sarmad@coie-nahrain.edu.iq

ISSN: 0975-7538

DOI: https://doi.org/10.26452/ijrps.v11i2.2115

Production and Hosted by

IJRPS | www.ijrps.com

INTRODUCTION

Learning Vector Quantization (LVQ) is a family of algorithms for statistical pattern classification, which aims at learning prototypes (codebook vectors) representing class regions (Nova and Estévez, 2014; Al-Faiz *et al.*, 2019a)

Backpropagation is another tool in artificial neural networks where its main purpose to sum up errors and learn the network about the dataset and accumulate errors and then update a map of weights in single or multiple layers between input and output in a supervised environment.

The nature of operation in this paper is input-output mapping, where each input must be mapped to a unique and independent output.

The dataset used in this paper was real gel electrophoresis (Calladine, 2004) images and 6 features where extracted according to (Anysz *et al.*, 2016).

Features extraction

The extraction of features was done according to (Anysz *et al.*, 2016), where 6 features were extracted from the gel electrophoresis images, as shown in Table 1,

Data set formulation steps

The gel electrophoresis images, as shown in Figure 1, must be saved in a previously defined loca-

Table 1: Features extracted from gel images

Feature name	Expression	Description
Avg	$f1 = \mu$	The average intensity in a region of an object
Std	$f2 = \sqrt{\sum_X (x - f1)\text{^}2\ H(X)}$	The standard deviation of intensity in a region of an object
Smoothness	$f3 = 1 - \frac{1}{(1+f1^2)}$	The relative smoothness of intensity in a region
Skewness	$f4 = \sum_X (x - f1)\text{^}3\ H(x)$	Deviation from the symmetry of mean intensity
Uniformity	$f5 = \sum_X H\text{^}2(X)$	Sum of a squared element in a histogram
Entropy	$f6 = -\log_2\ H(X)\ \sum_x H(X)$	A statistical measure of uncertainty

Table 2: comparison between LVQ and error-back-propagation

Method	Time of learning	Time of recognition	Iteration ruired to learn	Architecture
LVQ	313 Seconds	0.174 Seconds	308	384-769-6-6
BP	7.383 Seconds	0.116 Seconds	46	384-769-1

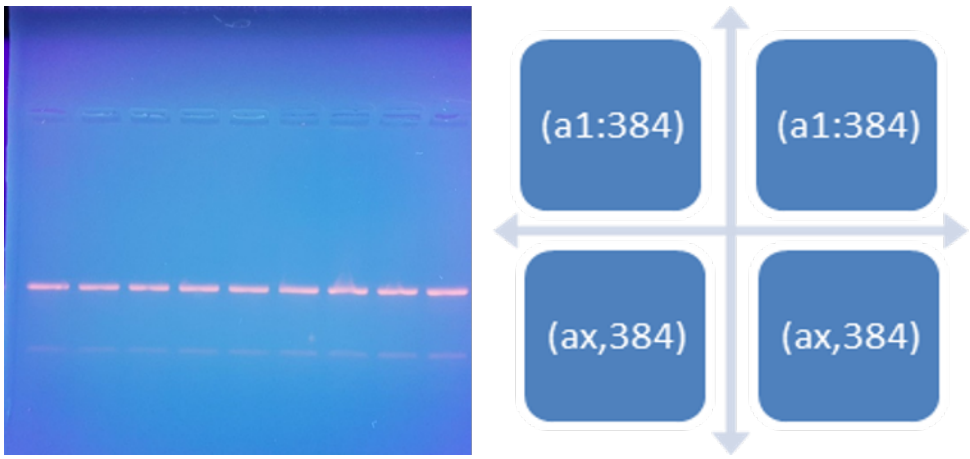

Figure 1: Gel electrophoreses image

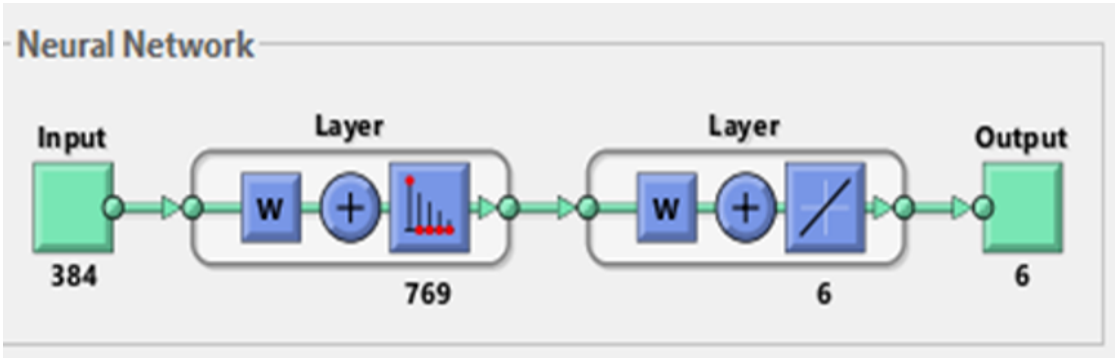

Figure 2: The LVQ network structure

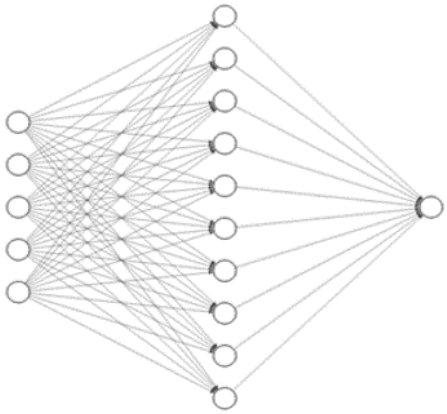

Figure 3: NN architecture

tion in the computer or the network. A MATLAB file reads the contents of the folder that contains the images and automatically loads each image. The features of each image were extracted and saved into a text file. After reading all images and saving a text file for each image, a single text file was created from all the text files. The features then converted to binary and then standardized using Z-score standardization (Al-Faiz *et al.*, 2019b), which can be calculated using Equation (1), as shown below,

$$Z(ij) = (a(ij) - \mu)/\sigma \quad (1)$$

Where,

Z(ij) is the new value

a(ij) is the old value

μ=the mean of the column of the input value

σ=the standard deviation of the column of the input value.

i=1— n (n=the number of rows or inputs)

j=1 —k (k=the number of bits that represents each input), then a MATLAB m file reads the contents of the folder that contains the images, and the features extracted from the equation, as shown above, where accumulated and saved together into a single text file.

Features were extracted from the gel electrophoresis images, which are,

Average standard deviation smoothness skewness uniformity entropy

Each feature then converted to a signed binary, which is represented by 64 binary bits, so each sample was represented by 384 bits, then a matrix of 384 columns and x rows was created, where x represents the number of gel electrophoresis images.

The input data had been structures as follows,

Where x inputs had been used, and each row in the above matrix represents a different input, and then each transposed row was fed into the neural network for learning.

Each row represents a single input, where each row is unique (because each row was generated from different gel electrophoresis images).

Proposed model overview

Two systems had been used for training and identification purposes: LVQ and Back-Propagation.

Linear Vector Quantization

The LVQ was used to train the neural network. The data-set used was the actual features extracted from gel electrophoresis images (the text file).

The LVQ network structure was 384-769-6-6 (384 input neurons - 769 neurons in the first hidden layer - 6 neurons in the second hidden layer - 6 neurons in the output layer), as shown below in Figure 2.

Error-Back-Propagation

(Whittington and Bogacz, 2019; Skorpil and Stastny, 2006). The structure of the Back-Propagation network used in this paper is as follows,

(384-769-1), 384 binary inputs (384 neurons)- 769 neurons in the hidden layer and one decimal output (one neuron) as shown in Figure 3.

Learning process

below is a flowchart of the learning process Figures 4, 5, 6 and 7,

The learning process is recursive and it consists of the following steps,

1. Each input (six features) must be fed to the system
2. The output of the system calculated

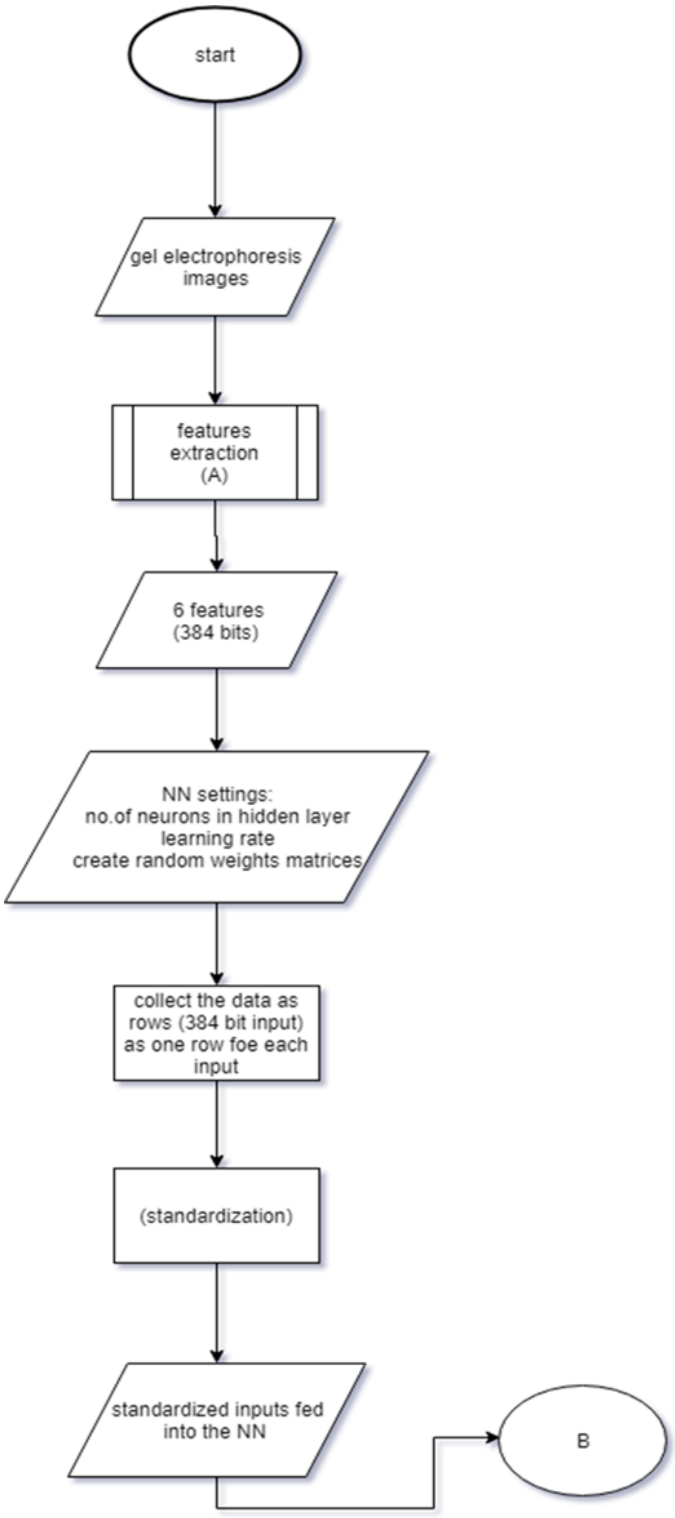

Figure 4: Learning process

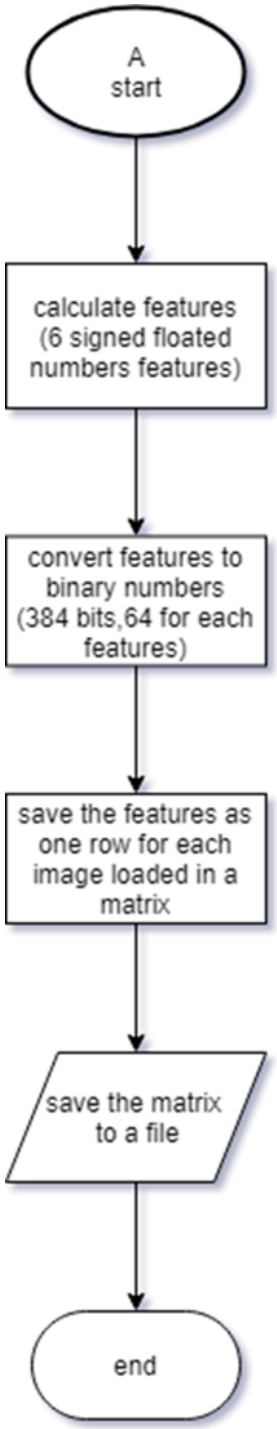

Figure 5: Learning process

3. The error is calculated according to equation 2, as shown below,
4. Desired output is a decimal number between (1-6) where each number represents a different input and that number is unique and will be used later for identification purposes.

$$a.\ error = desired\ output - actual\ output \quad (2)$$

5. The actual output is the summation of weighted inputs (between the hidden layer and the output neuron)
6. Then the second input was fed to the system (six feathers of the second image)
7. Another error was calculated according to Equation (2).
8. The errors were saved in an array

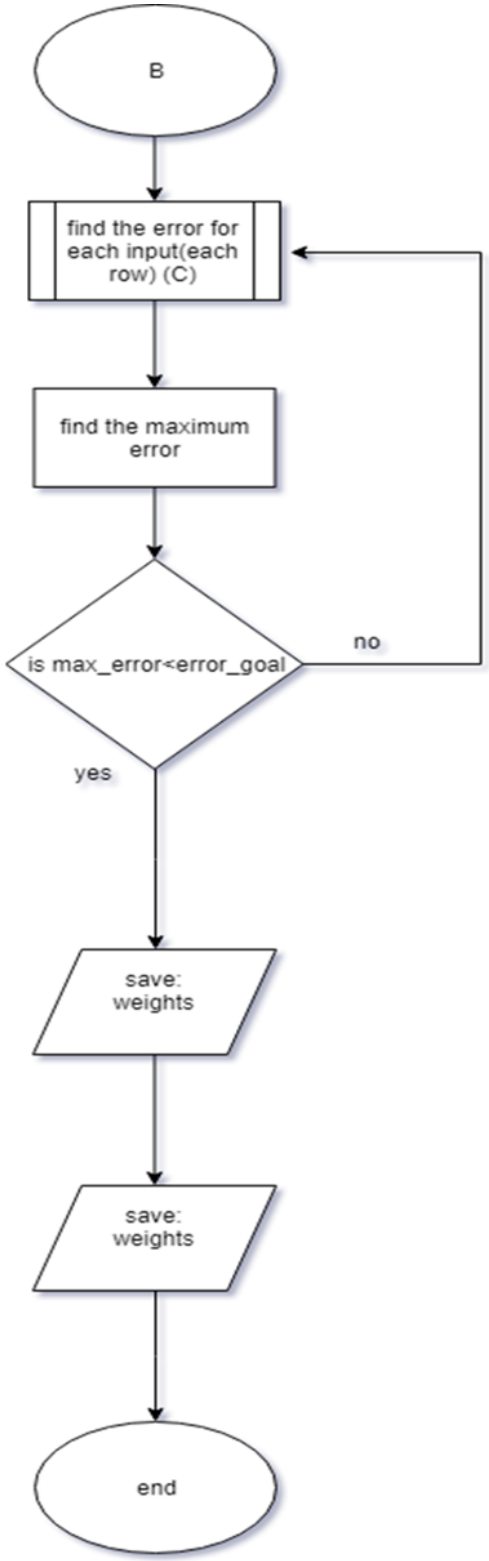

Figure 6: Learning process

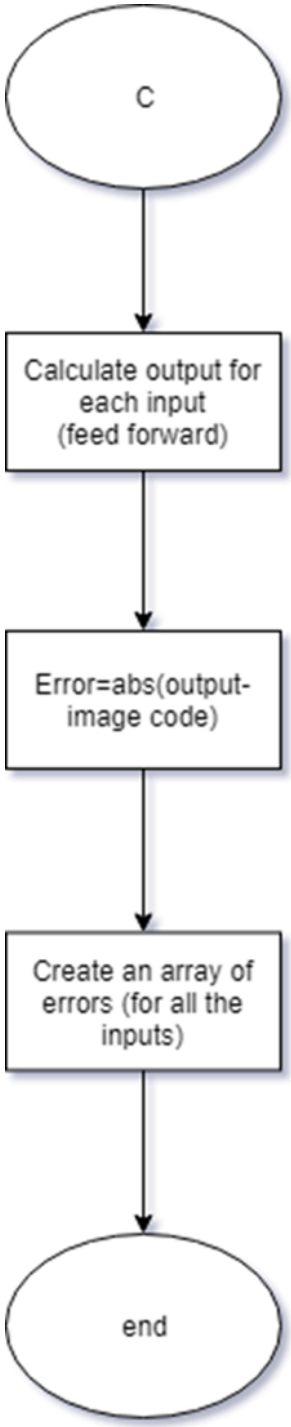

Figure 7: Learning process

9. The maximum error was calculated

10. A max error must be smaller than a threshold (10^{-4}), if not,

11. The all six images must be fed again to the network

12. Another max error is calculated

13. Compared again with the max error allowed

14. During this step above, the weights map must be updated accordingly until the goal is reached

15. That is the end of the learning process

16. The weights map must be saved in an external file.

The identification process is similar to the learning process as below,

1. load the gel-electrophoresis image to the system

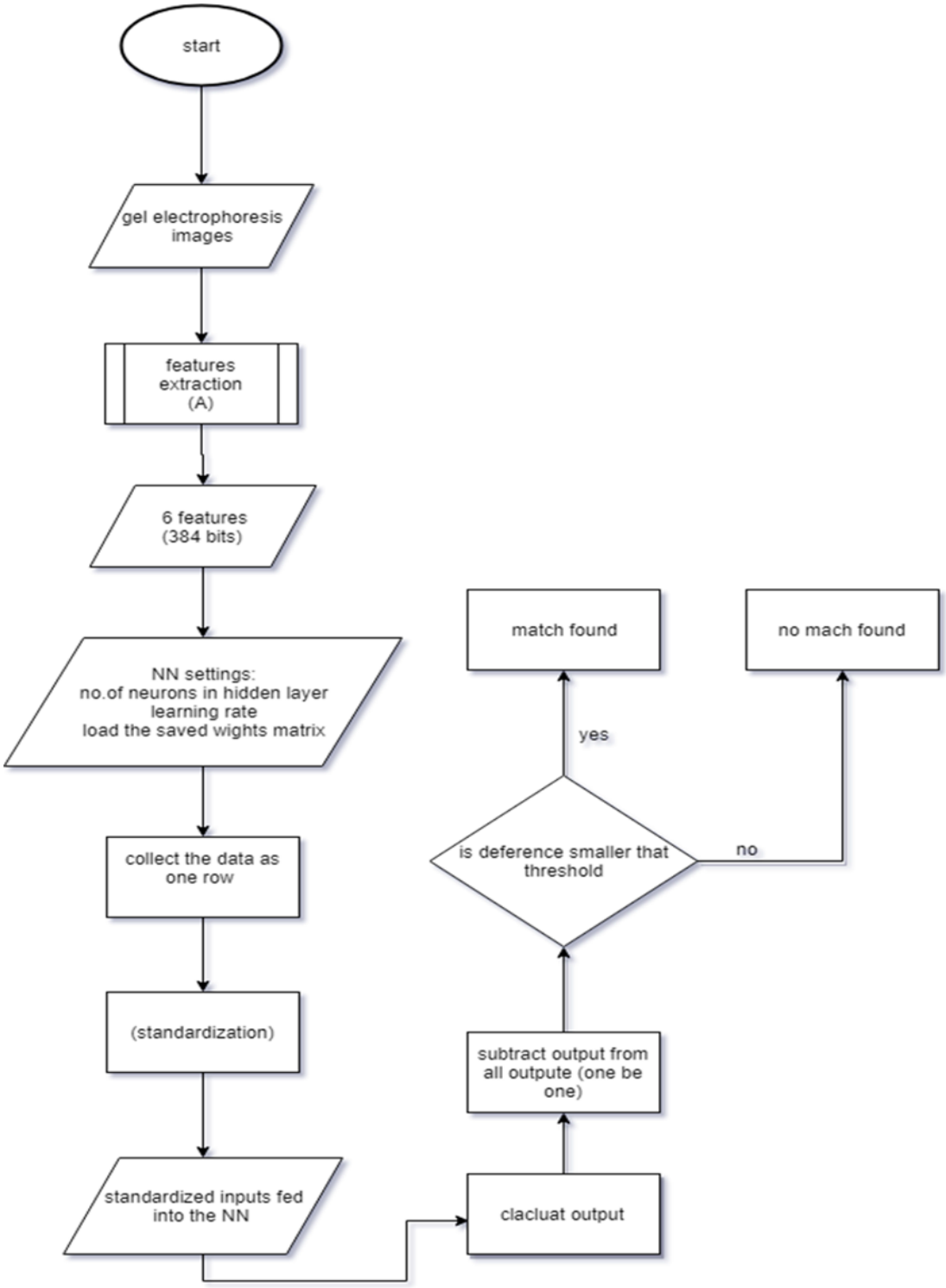

Figure 8: Identification process flowchart

2. extract the features from the image
3. save the features in an external file
4. load the weight matrix
5. load the features
6. standardize it
7. calculate the output of the network (forward only)
8. subtract the output from desired outputs (1-6)
9. when the difference is very low (10-4), the system declares an identification
10. if not, the system declares that "no match found"

A flowchart for the identification process is as shown below in Figure 8.

RESULTS AND DISCUSSION

After 296 LVQ Epoch, the system learned, then the system was tested by the learned data and it was successful 100% in identifying the learned data.

The LVQ is mainly used for classification of the input data, but in the proposed system, it was used for identification, but it was very slow in the learning process compared to error-back-propagation, as shown below in Table 2.

So, the error-back-propagation had been used as an alternative, and as shown above, in Table 1, it shows much better results in all terms.

CONCLUSIONS

From the results above, the following points had been concluded,

1-The obtained results show that LVQ can be used in identification in spite of the fact that LVQ's main job is classification.

2-The time of learning heavily depends on the number of inputs loaded into the system.

3-The standardization is a must because of the nature of the input data.

4-The proposed system can be used efficiently in forensics.

5- The proposed system can replace direct database queries efficiently because it does not have to search through all inputs (as in the normal SQL query).

6-Error-back-propagation is a better replacement for LVQ for the identification process.

REFERENCES

Al-Faiz, M. Z., Ibrahim, A. A., Hadi, S. 2019a. Gel Electrophoresis features extraction and pattern recognition using LVQ neural network. *7th International Student Congress on New Technologies in Engineering*.

Al-Faiz, M. Z., Ibrahim, A. A., Hadi, S. M. 2019b. The effect of Z-Score standardization (normalization) on binary input due to the speed of learning in a back-propagation neural network. *Iraqi Journal of Information & Communications Technology*, 1(3):42–48.

Anysz, H., Zbiciak, A., Ibadov, N. 2016. The Influence of Input Data Standardization Method on Prediction Accuracy of Artificial Neural Networks. *Procedia Engineering*, 153:66–70.

Calladine, C. R. 2004. Understanding DNA: The Molecule and How it Works: Third Edition, Understanding DNA: The Molecule and How it Works: Third Edition.

Nova, D., Estévez, P. A. 2014. A review of learning vector quantization classifiers. *Neural Computing and Applications*, pages 511–524.

Skorpil, V., Stastny, J. 2006. Neural Networks and Back Propagation Algorithm. *Electronics. Bulgaria, Sozopol*, pages 20–22.

Whittington, J. C. R., Bogacz, R. 2019. Theories of Error Back-Propagation in the Brain. *Trends in Cognitive Sciences. Elsevier Ltd*, 23(3):235–250.

